

Revealed Preferences over Risk and Uncertainty[†]

By MATTHEW POLISSON, JOHN K.-H. QUAH, AND LUDOVIC RENOU*

We develop a nonparametric method, called Generalized Restriction of Infinite Domains (GRID), for testing the consistency of budgetary choice data with models of choice under risk and under uncertainty. Our test can allow for risk-loving and elation-seeking attitudes, or it can require risk aversion. It can also be used to calculate, via Afriat's efficiency index, the magnitude of violations from a particular model. We evaluate the performance of various models under risk (expected utility, disappointment aversion, rank-dependent utility, and stochastically monotone utility) using data collected from several recent portfolio choice experiments. (JEL C14, D11, D12, D81)

This paper is a methodological contribution to the empirical investigation of decision making under risk and under uncertainty. While the expected utility (EU) model is the most widely used model for decision making in these contexts, there is also active research developing models that give a better account of observed choice behavior. A literature that tests the EU and other models on experimental data has emerged alongside these theoretical developments. These experiments often employ elicitation procedures in which subjects make repeated choices between two risky or uncertain outcomes; the data obtained in this way consist of a finite number of *binary choices*, which can then be used to partially recover a subject's preference. A more recent strand of experiments employs a different elicitation procedure, which we shall call the *budgetary choice* procedure, where subjects choose a preferred option from an effectively infinite set of alternatives. For example, a subject could

*Polisson: Department of Economics, University of Bristol, and Institute for Fiscal Studies (email: matthew.polisson@bristol.ac.uk); Quah: Department of Economics, Johns Hopkins University, and Department of Economics, National University of Singapore (email: john.quah@jhu.edu); Renou: School of Economics and Finance, Queen Mary University of London, and School of Economics, University of Adelaide (email: lrenou.econ@gmail.com). Stefano DellaVigna was the coeditor for this article. This research has made use of the ALICE High Performance Computing Facility at the University of Leicester. Part of this research was carried out while Matthew Polisson was visiting Caltech and UC Berkeley, and while Ludovic Renou was visiting the University of Queensland, and they would like to thank these institutions for their hospitality and support. Ludovic Renou would like to acknowledge financial support from the French National Research Agency (ANR), under the grant CIGNE (ANR-15-CE38-0007-01). We are grateful to five anonymous referees for insightful and constructive contributions which greatly improved the paper. We are also grateful to Ray Fisman, Yoram Halevy, Shachar Kariv, Felix Kübler, Lionel Page, Collin Raymond, and David Rojo Arjona for helpful discussions and comments, to Josh Lanier for superb RA support, as well as to seminar audiences at Adelaide, Bonn, Bristol, Brown, Caltech, Chapman (ESI), EUI, George Mason (ICES), Glasgow, IFN (Stockholm), Johns Hopkins, Lancaster, Leicester, NUS, NYU Stern, Oregon, Oxford, PSE, QUT, Queen Mary, Queensland, St. Andrews, Toulouse, UBC, UC Berkeley, UC Irvine, and UC San Diego, and to participants at many workshops and conferences.

[†]Go to <https://doi.org/10.1257/aer.20180210> to visit the article page for additional materials and author disclosure statements.

face a portfolio problem where she allocates her budget between two assets with state-contingent payoffs. An early experiment of this kind, the data from which we analyze in this paper, is found in Choi et al. (2007).¹ Other examples include Loomes (1991); Gneezy and Potters (1997); Bayer et al. (2013); Ahn et al. (2014); Choi et al. (2014); Hey and Pace (2014); Cappelen et al. (2015); and Halevy, Persitz, and Zrill (2018).

For reasons which we explain below, the nonparametric evaluation of data collected through a budgetary choice procedure requires a new approach. The contribution of this paper is twofold: (i) we develop a new empirical method that could be used to analyze data (be it experimental or field data) collected from portfolio decisions, and (ii) we apply this new method to evaluate the performance of different models of choice under risk using data from a number of recent portfolio choice experiments. Our method allows us to determine whether a dataset is consistent with the EU model or some of its generalizations, *without* making parametric assumptions on the Bernoulli function (such as constant relative risk aversion) or on other features of the model. This is empirically important because if we happen to find that a dataset is incompatible with a given model, then we can safely conclude that this incompatibility is attributable to the model itself rather than a poorly selected parametric form. Since the test also yields a utility function (which need not be unique) that best fits the data, this can be used to make out-of-sample predictions.

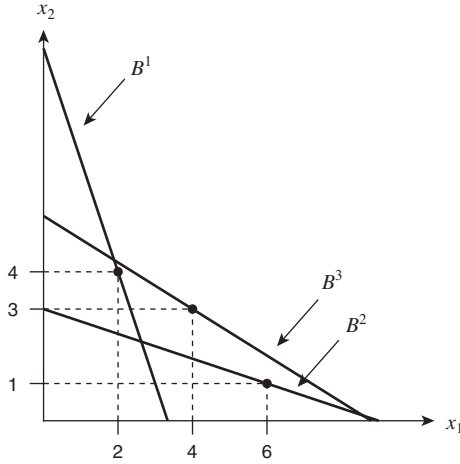
Our method can also be applied to test models of intertemporal choice (such as discounted utility) and other models which are formally similar to the EU model and its generalizations. Budgetary choice procedures are increasingly used in experiments to study intertemporal consumption (see, for example, Andreoni and Sprenger 2012 and Imai and Camerer 2018).

Testing EU and Other Models on a Finite Grid.—A feature of the budgetary choice procedure is that instead of requiring a subject to choose one alternative or another, it allows her to calibrate a response and to choose something “in between.” But this feature is also the crucial reason why the nonparametric analysis of data collected from this procedure requires a new empirical method, whereas no such method is necessary for binary choices. Indeed, suppose that we make a finite number of observations, where at observation t a subject chooses a lottery that gives a monetary payoff x_s^t in state s over one that gives y_s^t in state s (for $s = 1, 2, \dots, \bar{s}$), and where the probability of state s is commonly known to be $\pi_s > 0$. Imagine that we would like to test if this dataset is consistent with the EU model. Checking for exact consistency with the EU model simply involves finding a strictly increasing Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\sum_{s=1}^{\bar{s}} \pi_s u(x_s^t) \geq \sum_{s=1}^{\bar{s}} \pi_s u(y_s^t)$ holds at every observation t . This amounts to solving a finite set of linear inequalities,² and it is computationally straightforward to ascertain if a solution exists. However, it is clear that this method will no longer work when the subject is instead choosing from (classical) linear budget sets, since even a single observed choice from a budget set reveals an infinite set of binary preferences between the chosen bundle and alternatives in the budget.

¹ See this paper also for an account of the advantages of a budgetary choice approach.

² The unknowns to be solved for are $\{u(r) : r = x_s^t \text{ or } y_s^t \text{ for some } t \text{ and } s\}$.

Panel A. Dataset with three observations



Panel B. Dataset with grid inserted

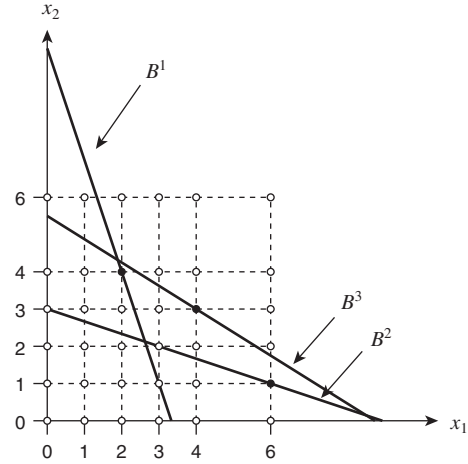


FIGURE 1. THE GRID METHOD

We now give a short and intuitive explanation of our new method. Consider a dataset with three observations and two states, as depicted in panel A of Figure 1; the horizontal axis corresponds to consumption in state 1, and the vertical axis to consumption in state 2. The subject chooses the contingent consumption bundle $\mathbf{x}^1 = (2, 4)$ from budget set B^1 , $\mathbf{x}^2 = (6, 1)$ from B^2 , and $\mathbf{x}^3 = (4, 3)$ from B^3 , where B^1 , B^2 , and B^3 are linear budget sets.³ Assume that the probability of state s is commonly known to be π_s . This dataset is said to be consistent with the EU model (or EU-rationalizable) if there is a continuous and strictly increasing Bernoulli function u such that $\pi_1 u(2) + \pi_2 u(4) \geq \pi_1 u(x_1) + \pi_2 u(x_2)$ for all $\mathbf{x} = (x_1, x_2)$ in B^1 , and similarly at the other two observations.

We show (in Theorem 1) that this dataset can be rationalized by the EU model if it can be rationalized on an appropriately restricted consumption set. Let \mathcal{X} be the set of consumption levels that are observed to have been chosen at some observation and in some state, plus zero; in this example $\mathcal{X} = \{0, 1, 2, 3, 4, 6\}$. Then for the dataset to be EU-rationalizable, it is *sufficient* (and obviously necessary) for it to be EU-rationalizable on the finite set \mathcal{X}^2 , i.e., there is a strictly increasing function $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ such that the expected utility of $\mathbf{x}^1 = (2, 4)$ is greater than any other bundle in $B^1 \cap \mathcal{X}^2$, and so forth.⁴ We refer to \mathcal{X}^2 as the *grid generated by the dataset*; it is depicted by the open circles in panel B of Figure 1. Therefore, checking if a dataset is EU-rationalizable involves checking if there is a solution to a finite set of linear inequalities, a problem which is computationally feasible to solve.

Since the key to our test is the restriction of an infinite consumption space to a finite grid, we refer to it as the method of Generalized Restriction of Infinite Domains

³Note that each B^i consists of the budget line and all of the bundles below the line.

⁴For example, since $(1, 6) \in B^1 \cap \mathcal{X}^2$, \bar{u} must satisfy $\pi_1 \bar{u}(2) + \pi_2 \bar{u}(4) > \pi_1 \bar{u}(1) + \pi_2 \bar{u}(6)$. A full set of inequalities that \bar{u} must satisfy is displayed in Table 1.

(GRID).⁵ The GRID method can also be used to test for consistency with other models of choice under risk (such as the rank-dependent utility (RDU) model (Quiggin 1982) and the disappointment aversion (DA) model (Gul 1991) and under uncertainty (such as the maxmin expected utility model (Gilboa and Schmeidler 1989)).⁶

The test we just described requires that the Bernoulli function be continuous and strictly increasing, but not necessarily concave, so a risk-loving EU-maximizer would pass the test. This is as it should be if model consistency is the principal concern. The stronger hypothesis that a subject is an EU-maximizer with a concave Bernoulli function can also be tested using a GRID method, where the test involves checking the superiority of the chosen bundle against a finite set of alternatives within each budget, though that finite set is constructed differently (and will no longer be the intersection of the budget with \mathcal{X}^2).

Empirical Implementation and Findings.—We implement our empirical method on three datasets obtained from the well-known portfolio choice experiments in Choi et al. (2007, 2014) and Halevy, Persitz, and Zrill (2018). In doing so, we are able to demonstrate the versatility and practicality of the GRID method, and also to reveal some empirical features common to all three datasets. In the Choi et al. (2007) experiment, each subject was asked to purchase Arrow-Debreu securities under different linear budget constraints. There were two states of the world, and it was commonly known that states occurred either symmetrically (each with probability 1/2) or asymmetrically (one with probability 1/3 and the other with probability 2/3); the experimental designs in Choi et al. (2014) and Halevy, Persitz, and Zrill (2018) closely resemble the symmetric design in Choi et al. (2007).

We use the GRID method developed in this paper to test the model performance of the EU, DA, and RDU models. We also check whether a subject's observations are consistent with the maximization of some locally nonsatiated utility function on the contingent consumption space. This is the most permissive utility model possible and forms the backdrop to our empirical analysis; Afriat's (1967) Theorem tells us that compatibility with utility maximization can be assessed by testing the Generalized Axiom of Revealed Preference (GARP).⁷ The GARP test could be strengthened to test for consistency with the maximization of a utility function that is *stochastically monotone*, in the sense that if a bundle dominates another with respect to first-order stochastic dominance, then it must have higher utility; a test for stochastically monotone utility maximization has recently been developed by Nishimura, Ok, and Quah (2017), and we implement it in this paper for the first time. The EU, DA, and RDU models are all special cases of stochastically monotone utility maximization, which is in turn more stringent than locally nonsatiated utility maximization.

In a rich budgetary choice environment with many observations on behavior, a dataset would typically not pass GARP (let alone more stringent requirements)

⁵It is "generalized" because one could think of the domain restriction used in this case (the grid \mathcal{X}^2) as a generalization of the one used in Afriat's Theorem (see Section IB).

⁶Our contribution should not be confused with Polisson and Quah (2013), which studies conditions for rationalizability in the presence of one or more goods that can only be consumed in discrete amounts. In that case, the consumption space is actually discrete; here, we show how to check for EU-rationalizability via a suitable discretization of the consumption space, but the space is not, in fact, discrete.

⁷This term and its acronym were coined by Varian (1982), which also provided a proof of Afriat's Theorem.

exactly. It is, however, possible to quantify a dataset's departure from rationalizability by a given model using the *critical cost efficiency index* (Afriat 1973); this index is widely used in the empirical revealed preference literature, including in Choi et al. (2007, 2014), while Halevy, Persitz, and Zrill (2018) implements a variant of this index first proposed by Varian (1990). The efficiency index runs from 1 to 0, with the index equal to 1 if a dataset passes the test exactly. We adopt this measure of rationality throughout our empirical implementation.⁸

For each subject, we are able to calculate that subject's efficiency indices for the different models under consideration. Across the three experiments, a negligible number of subjects pass GARP exactly. In the case of the primarily undergraduate subjects in Choi et al. (2007) and Halevy, Persitz, and Zrill (2018), more than 80 percent would pass GARP if we set a threshold of 0.9 for the efficiency index. In the case of the large-scale experiment (involving a representative sample) in Choi et al. (2014), the efficiency indices for GARP are distinctly lower, with nearly 60 percent passing GARP at the 0.9 efficiency threshold. The following highlights some salient features of the data collected from the three experiments:

- A significant minority of subjects either violate GARP and/or stochastic monotonicity; the decisions of these subjects cannot be explained by the EU, DA, or RDU models, all of which respect first-order stochastic dominance.
- Around one-half of the subjects who pass GARP (at some reasonable efficiency threshold) would also be compatible with the EU model; for these subjects, the EU model seems a good model of behavior, provided that some allowance is made for optimization errors.
- We find no evidence that the DA model accounts for the behavior of a significant proportion of subjects not accounted for by the EU model.
- On the other hand, there is some evidence that the RDU model could explain a significant segment of the population not behaving as EU-maximizers.⁹

Since our testing procedure also produces, for each subject, a rationalizing utility function belonging to a given model, that recovered utility function could then be used to make out-of-equilibrium predictions. We carry out a simple exercise of this type, using a rank-dependent utility function estimated from a subject's portfolio decisions, to make predictions on the subject's choice when she is independently presented with a choice between two lotteries. Our objective is to not check how often the procedure makes correct predictions, since the data we have access to do not allow us to explore that question in a meaningful way, but simply to illustrate the potential usefulness of our nonparametric methods for this purpose.

Relationship with the Revealed Preference Literature.—Our paper is related to the revealed preference literature originating from Afriat's (1967) Theorem, which characterizes price and demand observations that are consistent with the maximization

⁸We also carry out some empirical analysis using Varian's index, which is reported in the online Appendix. The calculation of Varian's index is more computationally demanding than calculating Afriat's, so our analysis using that index does not cover all of the models under consideration.

⁹When there are two equiprobable states, the RDU and DA models are known to be equivalent. Our finding that RDU explains significantly more behavior than EU is specific to the asymmetric treatment in Choi et al. (2007).

of a locally nonsatiated utility function (see also Diewert 1973 and Varian 1982). A natural follow-up to Afriat's contribution is to characterize those datasets which are rationalizable by more specialized utility functions. Among these papers are those which characterize state price and contingent consumption demand observations that are consistent with the EU model¹⁰ and (in more recent papers) some of its generalizations: these include Varian (1983a, b, 1988); Green and Srivastava (1986); Diewert (2012); Bayer et al. (2013); Echenique and Saito (2015); Chambers, Liu, and Martinez (2016); and Chambers, Echenique, and Saito (2016).¹¹ The principal difference between our results and this literature is that we do not rely on the sufficiency of first-order conditions. This has two important implications: (i) the models we consider need not induce a convex preference over the contingent consumption space (e.g., we allow for risk-loving behavior under EU or elation-seeking behavior under DA), and (ii) we can weaken the requirement that the constraint set is a linear budget set. For reasons which we give in Section II, allowing for nonlinear constraint sets enables our method to be used to calculate Afriat's efficiency index.

Organization of the Paper.—Section I describes how one can use the GRID method to test the EU, DA, and RDU models. When a dataset is not fully consistent with a particular model, the method can also help calculate the Afriat and Varian efficiency indices; this is described in Section II. Section III explains how the method could be extended to evaluate models where concavity is imposed on the Bernoulli function. The empirical implementation is in Section IV. Further applications of the GRID method to test models of decision making under risk/uncertainty, or over time, are found in the online Appendix. The online Appendix also describes and implements a new algorithm for calculating Varian's index for the locally nonsatiated (GARP), stochastically monotone, and expected utility models; readers interested in calculating Varian's index may find this useful, whether or not they plan to apply the GRID method.

I. The GRID Method

We assume that there is a finite set of states, denoted by $S = \{1, 2, \dots, \bar{s}\}$. The contingent consumption space is $\mathbb{R}_+^{\bar{s}}$; for a typical consumption bundle $\mathbf{x} \in \mathbb{R}_+^{\bar{s}}$, the s th entry, x_s , specifies the consumption level in state s .¹² There is a finite dataset

$$\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$$

¹⁰The EU model requires the utility function on the contingent consumption space to be additively separable. There are also results which characterize datasets that are rationalizable by a *weakly* separable utility function (see Varian 1983a and Quah 2014).

¹¹There is also a closely related literature on recovering expected utility from asset or contingent consumption demand functions, where, in effect the dataset is assumed to be infinite (see, for example, Dybvig and Polemarchakis 1981 and Kubler, Selden, and Wei 2014).

¹²Our results do depend on the realization in each state being one-dimensional (which can be interpreted as a monetary payoff, but not a bundle of goods). This case is the one most often considered in applications and experiments and is also the assumption in a number of recent papers, including Kubler, Selden, and Wei (2014); Echenique and Saito (2015); and Chambers, Echenique, and Saito (2016). The papers by Varian (1983a, b); Green and Srivastava (1986); Bayer et al. (2013); and Chambers, Liu, and Martinez (2016) allow for multidimensional realizations. However, and crucially, in all of these papers, convexity of the agent's preference over contingent consumption and linear budget sets are also required.

consisting of T observations, where $\mathbf{x}^t \in B^t$ and $B^t \subset \mathbb{R}_+^{\bar{s}}$. We could interpret this as data collected from an experiment where the subject chooses the bundle \mathbf{x}^t from the constraint set B^t at observation t and this will indeed be our interpretation throughout the paper.¹³ We assume that B^t is a compact set.

We denote the *upper boundary* of B^t by ∂B^t ; an element $\mathbf{x} \in B^t$ is said to be in ∂B^t if there is no $\mathbf{x}' \in B^t$ such that $\mathbf{x}' > \mathbf{x}$.¹⁴ The *downward extension* of B^t is the set

$$\underline{B}^t = \{\mathbf{y} \in \mathbb{R}_+^{\bar{s}} : \mathbf{y} \leq \mathbf{x} \text{ for some } \mathbf{x} \in B^t\}.$$

Obviously, \underline{B}^t contains B^t . The most important example of a constraint set is the *classical* or *linear budget set*. At price vector $\mathbf{p} \in \mathbb{R}_{++}^{\bar{s}}$ and wealth $w > 0$, the classical budget set is $\mathcal{B}(\mathbf{p}, w) = \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p} \cdot \mathbf{x} \leq w\}$. By a *classical dataset*, which we denote by $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$, we mean a dataset where, at observation t , the subject chooses \mathbf{x}^t from

$$(1) \quad B^t = \mathcal{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t) = \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p}^t \cdot \mathbf{x} \leq \mathbf{p}^t \cdot \mathbf{x}^t\}.$$

Thus, \mathcal{O} can also be written as $\{(\mathbf{x}^t, \mathcal{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t))\}_{t=1}^T$. Note that the upper boundary of $\mathcal{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t)$ is simply the *budget plane*, i.e., $\partial \mathcal{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t) = \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p}^t \cdot \mathbf{x} = \mathbf{p}^t \cdot \mathbf{x}^t\}$, while the downward extension of $\mathcal{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t)$ is itself. The experiments conducted in Choi et al. (2007, 2014) and Halevy, Persitz, and Zrill (2018), the data from which we analyze in Section IV, involve subjects choosing from classical budget sets with two states.

Bear in mind, however, that our formulation only requires B^t to be compact and it does *not* have to be a linear budget set. A nonlinear budget set occurs when a subject chooses contingent consumption through a portfolio of securities in an incomplete market (i.e., loosely speaking, when the number of securities is fewer than the number of states); in this case, the budget set will not be linear, but it will be compact so long as the security prices do not admit arbitrage.¹⁵ In this paper, the crucial application requiring B^t to be nonlinear is in *quantifying approximate rationalizability* when a dataset \mathcal{O} cannot be exactly rationalized; as we explain in Section II, this requires testing the rationalizability of a modified dataset that has nonlinear constraint sets, even if the true constraint sets are linear budget sets.

Before presenting the formal results, we provide an informal explanation of the general approach we adopt in ascertaining whether a dataset \mathcal{O} is compatible with a given model of decision making under risk or uncertainty. We first notice that most of these models have two essential components: a *Bernoulli function* $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and an *aggregator function* $\phi : \mathbb{R}_+^{\bar{s}} \rightarrow \mathbb{R}$, so that the utility of a bundle \mathbf{x} is $\phi(u(x_1), \dots, u(x_{\bar{s}}))$. For a *given* aggregator ϕ , the first step is to test whether there exists a Bernoulli function u that rationalizes the data, i.e., at each observation t ,

¹³ Obviously nothing in principle forecloses the possibility of applying our method to observational budgetary choice data of the type found in insurance or financial decision problems.

¹⁴ For the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{\bar{s}}$, we write $\mathbf{x} \geq \mathbf{y}$ if $x_s \geq y_s$ for all s , and $\mathbf{x} > \mathbf{y}$ if $\mathbf{x} \geq \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$; if $x_s > y_s$ for all s , we write $\mathbf{x} \gg \mathbf{y}$.

¹⁵ Indeed, there is $\mathbf{p}^t \gg 0$ such that $B^t = \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p}^t \cdot \mathbf{x} \leq \mathbf{p}^t \cdot \mathbf{x}^t\} \cap \{Z^t + \omega^t\}$, where Z^t is the span of assets available to the agent and ω^t is the agent's endowment of contingent consumption at observation t . Both B^t and \mathbf{x}^t will be known to the observer, if he knows the asset prices, the agent's holding of securities, the asset payoffs in every state, and the agent's endowment of contingent consumption.

the utility of the chosen bundle \mathbf{x}^t is weakly greater than the utility of any other bundle $\mathbf{x} \in B^t$. Theorem 1 provides that test. However, a model may correspond to a *family* of aggregators ϕ . If so, there is a second step that involves testing whether there is an aggregator in the family of aggregators under consideration *and* a Bernoulli function u which together rationalize the data. We now turn to the formal exposition.

Let $\{\phi(\cdot, t)\}_{t=1}^T$ be a collection of functions, where $\phi(\cdot, t) : \mathbb{R}_+^{\bar{s}} \rightarrow \mathbb{R}$ is continuous and strictly increasing.¹⁶ The dataset $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is said to be *rationalizable* by $\{\phi(\cdot, t)\}_{t=1}^T$ if there exists a continuous and strictly increasing function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, which we shall refer to as the *Bernoulli function*, such that

$$(2) \quad \phi(\mathbf{u}(\mathbf{x}^t), t) \geq \phi(\mathbf{u}(\mathbf{x}), t) \quad \text{for all } \mathbf{x} \in B^t,$$

where $\mathbf{u}(\mathbf{x}) = (u(x_1), u(x_2), \dots, u(x_{\bar{s}}))$. In other words, the observed choice behavior is consistent with the hypothesis that, at observation t , the subject has chosen a bundle from B^t that maximizes the utility function $\phi(\mathbf{u}(\cdot), t) : \mathbb{R}_+^{\bar{s}} \rightarrow \mathbb{R}$.¹⁷

The function $\phi(\cdot, t)$ aggregates the vector of “utils” $\mathbf{u}(\mathbf{x})$ into a single number. Of course, the most familiar formula for $\phi(\cdot, t)$ arises in the *expected utility* (EU) model; in this case, if the probability of state s at observation t is objectively known to be $\pi_s^t > 0$, then

$$(3) \quad \phi(u_1, u_2, \dots, u_{\bar{s}}, t) = \sum_{s=1}^{\bar{s}} \pi_s^t u_s.$$

Other models will lead to different formulations of ϕ (as we illustrate in Section ID). Note that the objective probabilities need not vary across observations; if they do not, ϕ would be independent of t .

Two requirements are imposed on the Bernoulli function u . Continuity is an important technical condition because it guarantees that $\phi(\mathbf{u}(\cdot), t)$ is continuous, which in turn guarantees that the agent’s utility maximization problem always has a well-behaved solution on compact constraint sets.¹⁸ The other requirement on u is that it is strictly increasing. Notice that *some* assumption of this type is necessary: in particular, if we allow u to be a constant function then *every* dataset \mathcal{O} is rationalizable because the subject would be indifferent across all bundles in $\mathbb{R}_+^{\bar{s}}$. Requiring u to be strictly increasing is reasonable since its argument is typically interpreted as money. This assumption, together with the assumption that $\phi(\cdot, t)$ is a strictly increasing function guarantees that

$$(4) \quad \phi(\mathbf{u}(\mathbf{x}^t), t) > \phi(\mathbf{u}(\mathbf{x}), t) \quad \text{for all } \mathbf{x} \in \underline{B}^t \setminus \partial B^t.¹⁹$$

¹⁶By strictly increasing, we mean that $\phi(\mathbf{z}, t) > \phi(\mathbf{z}', t)$ if $\mathbf{z} > \mathbf{z}'$.

¹⁷In keeping with the more empirically oriented parts of the revealed preference literature, this definition allows for the possibility that there are other bundles \mathbf{x} in B^t that maximize $\phi(\mathbf{u}(\cdot), t)$.

¹⁸To be precise, it guarantees that the optimal solutions form a nonempty compact set and is (in the case of demand) an upper hemicontinuous correspondence of prices.

¹⁹Indeed, if $\mathbf{x} \in \underline{B}^t \setminus \partial B^t$, then there is $\mathbf{y} \in B^t$ such that $\mathbf{y} \geq \mathbf{x}$ and by the optimality of \mathbf{x}^t , $\phi(\mathbf{u}(\mathbf{x}^t), t) \geq \phi(\mathbf{u}(\mathbf{y}), t)$. If $\mathbf{y} = \mathbf{x}$ then $\mathbf{y} \in B^t \setminus \partial B^t$, so $\phi(\mathbf{u}(\mathbf{x}^t), t) > \phi(\mathbf{u}(\mathbf{y}), t) = \phi(\mathbf{u}(\mathbf{x}), t)$. Otherwise, $\mathbf{y} > \mathbf{x}$ and by the strict increasing property, $\phi(\mathbf{u}(\mathbf{x}^t), t) \geq \phi(\mathbf{u}(\mathbf{y}), t) > \phi(\mathbf{u}(\mathbf{x}), t)$.

In other words, a bundle that is not on the boundary of the constraint set has strictly lower utility than the chosen bundle \mathbf{x}^t .

Note that, for three related reasons, we do not require u to be concave. First, the concavity of u is *not* a fundamental part of EU theory or many of its generalizations and it is desirable to have a test with a minimum of ancillary assumptions, so that any rejection of the model would be decisive and could not be attributed to the effect of those assumptions. This is relevant because we know that the concavity of u imposes observable restrictions on portfolio choices over and above those implied by the EU model (see online Appendix Section A4.1).²⁰ Second, in portfolio choice settings, concavity of u is often imposed because (along with other assumptions on ϕ) it ensures the convexity of the agent's preference over contingent consumption bundles (equivalently, the quasiconcavity of the utility function defined on \mathbb{R}_+^S); this in turn facilitates the mathematical analysis of portfolio choice. However, as emphasized by Halevy, Persitz, and Zrill (2017), there is a distinction between out-of-sample predictions made with and without the convexity property, and the sharper conclusions obtained by imposing this property can be misleading. Lastly, departures from the concavity of u have been specifically exploited to explain certain specific empirical phenomena; an early paper of that type is Friedman and Savage (1948). In prospect theory, the nonconcavity of the Bernoulli function around a reference point also plays a crucial role.²¹ So there is advantage in having a test that is agnostic about the curvature of u .

A. The Main Result

Let Y be any subset of \mathbb{R}_+ . Given a Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, the function $\bar{u} : Y \rightarrow \mathbb{R}_+$ is the *restriction* of u to Y , if the functions agree on Y , i.e., $\bar{u}(r) = u(r)$ for all $r \in Y$. In the other direction, a Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is said to *extend* a function $\bar{u} : Y \rightarrow \mathbb{R}_+$ if the two functions agree on Y .

Given a dataset $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$, we define

$$(5) \quad \mathcal{X}^* = \{0\} \cup \{r \in \mathbb{R}_+ : r = x_s^t \text{ for some } t \text{ and } s\};$$

besides 0, \mathcal{X}^* contains those levels of consumption that are chosen at some observation and in some state. Since the dataset is finite, so is \mathcal{X}^* . Let \mathcal{X} be a finite subset of \mathbb{R}_+ containing \mathcal{X}^* . We define $\mathcal{G} = \mathcal{X}^S$ and shall refer to \mathcal{G} as the *grid* associated with \mathcal{O} . Suppose that $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is rationalizable by $\{\phi(\cdot, t)\}_{t=1}^T$ with the Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. If $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ is the restriction of u to \mathcal{X} , then

$$(6) \quad \phi(\bar{\mathbf{u}}(\mathbf{x}^t), t) \geq \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \quad \text{for all } \mathbf{x} \in \underline{B}^t \cap \mathcal{G}$$

(where $\bar{\mathbf{u}}(\mathbf{x}) = (\bar{u}(x_1), \bar{u}(x_2), \dots, \bar{u}(x_S))$), and

²⁰Readers familiar with Afriat's Theorem will know that in that context, requiring the concavity of the rationalizing utility function U (defined on the consumption space \mathbb{R}_+^S) imposes no observable restrictions. However, it does not logically follow that imposing concavity on u is also innocuous, because the EU model implies that U must also be additive across states.

²¹An extension to gains/losses around a reference point is in online Appendix Section A3.4.

$$(7) \quad \phi(\bar{\mathbf{u}}(\mathbf{x}^t), t) > \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \quad \text{for all } \mathbf{x} \in (\underline{B}^t \setminus \partial B^t) \cap \mathcal{G}.$$

This follows immediately from (2) and (4) since $u(r) = \bar{u}(r)$ for all $r \in \mathcal{X}$ and $B^t \cap \mathcal{G} \subset B^t$ and $(\underline{B}^t \setminus \partial B^t) \cap \mathcal{G} \subset \underline{B}^t \setminus \partial B^t$. In other words, if u rationalizes the dataset \mathcal{O} then it will continue to rationalize the dataset if the consumption space is restricted to the grid \mathcal{G} . Our main theorem says that the converse of this statement is also true.²²

THEOREM 1: *Suppose that $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is rationalizable by the collection of continuous and strictly increasing functions $\{\phi(\cdot, t)\}_{t=1}^T$ with the Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Let \mathcal{X} be a finite set in \mathbb{R}_+ that contains \mathcal{X}^* (as defined by (5)) and let $\mathcal{G} = \mathcal{X}^s$. Then the restriction of u to \mathcal{X} , $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$, satisfies conditions (6) and (7).*

Conversely, suppose that given $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ and a collection of continuous and strictly increasing functions $\{\phi(\cdot, t)\}_{t=1}^T$, there is a strictly increasing function $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ that satisfies conditions (6) and (7). Then there is a Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that extends \bar{u} and with which $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is rationalizable by $\{\phi(\cdot, t)\}_{t=1}^T$.

This theorem tells us that testing for the rationalizability of \mathcal{O} is equivalent to testing for rationalizability in the case where the agent's consumption space is restricted to the grid \mathcal{G} . Crucially, this reduces the rationality requirements to a *finite* number of optimality conditions involving the observed choices and alternatives (see (6) and (7)). We refer to this approach to revealed preference testing as the method of Generalized Restriction of Infinite Domains (GRID), or simply, the GRID method or GRID test. The domain restriction is “generalized” because Afriat's Theorem could also be understood as providing a test with such an approach (see Section IB for more details), even though the domain restriction in the two results differ. Indeed there are results characterizing rationalizability in other models that could be thought of as using a GRID approach.²³

The intuition for Theorem 1 ought to be strong. Given \bar{u} satisfying (6) and (7), we can define the step function $\hat{u} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ where $\hat{u}(r) = \bar{u}([r])$, with $[r]$ being the largest element of \mathcal{X} weakly lower than r , i.e., $[r] = \max\{r' \in \mathcal{X} : r' \leq r\}$. Notice that $\phi(\hat{\mathbf{u}}(\mathbf{x}^t), t) = \phi(\bar{\mathbf{u}}(\mathbf{x}^t), t)$ and, for any $\mathbf{x} \in \underline{B}^t$, $\phi(\hat{\mathbf{u}}(\mathbf{x}), t) = \phi(\bar{\mathbf{u}}([\mathbf{x}]), t)$, where $[\mathbf{x}] = ([x_1], [x_2], \dots, [x_s])$ is in $\underline{B}^t \cap \mathcal{G}$. Clearly, if \bar{u} obeys (6) and (7) then \mathcal{O} is rationalized by $\{\phi(\cdot, t)\}_{t=1}^T$ and \hat{u} (in the sense that (2) holds). This falls short of

²²We cannot replace \underline{B}^t with B^t in (6) and (7). For example, suppose $\mathbf{x}^1 = (1, 0)$ is chosen from $B^1 = \{(x_1, x_2) \in \mathbb{R}_+^2 : 2x_1 + x_2 = 2\}$ and $\mathbf{x}^2 = (0, 1)$ is chosen from $B^2 = \{(x_1, x_2) \in \mathbb{R}_+^2 : x_1 + 2x_2 = 2\}$ (so the constraint sets are straight lines). These observations cannot be rationalized by any increasing utility function and, in particular, cannot be rationalized in the sense of Theorem 1 (with ϕ constant across t). However, since $\mathcal{G} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, $B^1 \cap \mathcal{G} = \{(1, 0)\}$ and $B^2 \cap \mathcal{G} = \{(0, 1)\}$, conditions (6) and (7) hold if \underline{B}^t is replaced with B^t . On the other hand $(\underline{B}^1 \setminus \partial B^1) \cap \mathcal{G}$ contains $(0, 1)$ and $(\underline{B}^2 \setminus \partial B^2) \cap \mathcal{G}$ contains $(1, 0)$, so (7) requires $\phi(\bar{\mathbf{u}}(\mathbf{x}^1)) > \phi(\bar{\mathbf{u}}(\mathbf{x}^2))$ and $\phi(\bar{\mathbf{u}}(\mathbf{x}^1)) < \phi(\bar{\mathbf{u}}(\mathbf{x}^2))$, which plainly cannot happen. This allows us to conclude, correctly, that this dataset is not rationalizable.

²³Proposition 1 devises a GRID test for EU-rationalizability with concave Bernoulli functions. Other papers using what could be broadly understood as a GRID approach include Quah (2014); Dziegowski and Quah (2014); and Chambers, Liu, and Martinez (2016). Note, however, that these results characterize different models, involve different domain restrictions, and have different proofs from Theorem 1.

the claim in the theorem only because \hat{u} is neither continuous nor strictly increasing;²⁴ the proof in the Appendix shows how one could in fact construct a function with these additional properties.

Note that Theorem 1 gives some leeway on how \mathcal{X} is chosen. If we are simply interested in testing for the rationalizability of \mathcal{O} by a given model, then we could pick $\mathcal{X} = \mathcal{X}^*$, but sometimes it is advantageous to let \mathcal{X} be a strictly larger set (see Section IC on making out-of-sample predictions). Note also that in checking the conditions (6) and (7) we can confine ourselves to checking those bundles \mathbf{x}' in $\underline{B}^t \cap \mathcal{G}$ which are not dominated by some other bundle in $\underline{B}^t \cap \mathcal{G}$. This is because if $\mathbf{x}' > \mathbf{x}''$ and property (6) or (7) holds for $\mathbf{x} = \mathbf{x}'$, it will also hold for $\mathbf{x} = \mathbf{x}''$ since both ϕ and \bar{u} are strictly increasing.

B. Testing the Expected Utility Model

Theorem 1 provides us with a convenient way of testing for rationalizability by the EU model. Consider an experiment where the probability of any state can be (possibly) varied across observations and where these probabilities are announced, so that both the observer and the subject know that the probability of state s at observation t is $\pi_s^t > 0$. The dataset $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is *EU-rationalizable* if there is a Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\sum_{s=1}^{\bar{s}} \pi_s^t u(x_s^t) \geq \sum_{s=1}^{\bar{s}} \pi_s^t u(x_s)$ for all $\mathbf{x} \in B^t$. Theorem 1 tells us that $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is EU-rationalizable if and only if there is a strictly increasing function $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ for which (6) and (7) hold, with ϕ given by (3), i.e.,

$$(8) \quad \sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s^t) \geq \sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s) \quad \text{for all } \mathbf{x} \in \underline{B}^t \cap \mathcal{G}$$

and

$$(9) \quad \sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s^t) > \sum_{s=1}^{\bar{s}} \pi_s^t \bar{u}(x_s) \quad \text{for all } \mathbf{x} \in (\underline{B}^t \setminus \partial B^t) \cap \mathcal{G}.$$

This is a system of linear inequalities, and solving it is both formally possible (in the sense that there is an algorithm that can decide within a known number of steps whether it has a solution) and computationally feasible.

As an example of how this works in practice, consider again the dataset depicted in panel A of Figure 1. Suppose that it is commonly known that the probability of state s ($s = 1, 2$) at observation t is π_s^t . Since the three observed choices are $\mathbf{x}^1 = (2, 4)$, $\mathbf{x}^2 = (6, 1)$, and $\mathbf{x}^3 = (4, 3)$, $\mathcal{X}^* = \{0, 1, 2, 3, 4, 6\}$. Choosing $\mathcal{X} = \mathcal{X}^*$, EU-rationalizability can be tested by checking for a solution to the conditions listed in Table 1. In the top-left panel are the strict inequalities guaranteeing that \bar{u} is strictly increasing. The other panels list the conditions for the optimality of \mathbf{x}^1 in B^1 , \mathbf{x}^2 in B^2 , and \mathbf{x}^3 in B^3 .²⁵ For example, at observation 1, the observed choice is $\mathbf{x}^1 = (2, 4)$ and there are 18 bundles in $B^1 \cap \mathcal{G}$ (besides \mathbf{x}^1), of which only the 2 bundles $(1, 6)$

²⁴ Recall that the Bernoulli function u is continuous and strictly increasing by definition.

²⁵ Recall that since each B^t is a linear budget set, we have $\underline{B}^t = B^t$ for all t .

TABLE 1—CONDITIONS ON \bar{u} FOR EU-RATIONALIZABILITY

| Monotonicity of \bar{u} | Optimality of $\mathbf{x}^1 = (2, 4)$ |
|--|--|
| $\bar{u}(6) > \bar{u}(4) > \bar{u}(3) > \bar{u}(2)$ | $\pi_1^1 \bar{u}(2) + \pi_2^1 \bar{u}(4) > \pi_1^1 \bar{u}(1) + \pi_2^1 \bar{u}(6)$ |
| $\bar{u}(2) > \bar{u}(1) > \bar{u}(0)$ | $\pi_1^1 \bar{u}(2) + \pi_2^1 \bar{u}(4) \geq \pi_1^1 \bar{u}(3) + \pi_2^1 \bar{u}(1)$ |
| Optimality of $\mathbf{x}^2 = (6, 1)$ | Optimality of $\mathbf{x}^3 = (4, 3)$ |
| $\pi_1^2 \bar{u}(6) + \pi_2^2 \bar{u}(1) \geq \pi_1^2 \bar{u}(0) + \pi_2^2 \bar{u}(3)$ | $\pi_1^3 \bar{u}(4) + \pi_2^3 \bar{u}(3) > \pi_1^3 \bar{u}(2) + \pi_2^3 \bar{u}(4)$ |
| $\pi_1^2 \bar{u}(6) + \pi_2^2 \bar{u}(1) \geq \pi_1^2 \bar{u}(3) + \pi_2^2 \bar{u}(2)$ | $\pi_1^3 \bar{u}(4) + \pi_2^3 \bar{u}(3) > \pi_1^3 \bar{u}(6) + \pi_2^3 \bar{u}(1)$ |

and $(3, 1)$ are undominated,²⁶ with the former in the interior of the budget set and the latter on the upper boundary. The first inequality in the top-right panel is imposed by (9) and the second inequality by (8). Similarly, the reader can check that there are two undominated bundles in $B^2 \cap \mathcal{G}$ and two in $B^3 \cap \mathcal{G}$, leading to the inequality conditions displayed in the bottom panels. EU-rationalizability holds if there is a \bar{u} that solves the linear inequalities displayed in Table 1.²⁷

At this point it is worth emphasizing that requiring a dataset to be EU-rationalizable is certainly more stringent than simply requiring it to be rationalizable by a locally nonsatiated utility function (on the contingent consumption space \mathbb{R}_+^5). Indeed, while a dataset with a single observation $(\mathbf{x}^1, \mathbf{p}^1)$ must necessarily be rationalizable in that sense, even a single observation can be incompatible with the EU model.

Example 1: Suppose that at the price vector $\mathbf{p}^1 = (1, 2)$, the subject chooses the bundle $\mathbf{x}^1 = (1, 2)$. This subject is buying more of the more expensive good, which is incompatible with the maximization of expected utility when the two states are equiprobable. It would, of course, fail the GRID test. Indeed, let $\mathcal{X} = \mathcal{X}^* = \{0, 1, 2\}$. In panel A of Figure 2, we depict \mathbf{x}^1 chosen from $B^1 = \underline{B}^1 = \{\mathbf{x} \in \mathbb{R}_+^2 : x_1 + 2x_2 \leq 5\}$, and with the grid $\mathcal{G} = \mathcal{X}^2$ inserted in panel B. Clearly, $(2, 1) \in \mathcal{G} \cap (B^1 \setminus \partial B^1)$; comparing $\mathbf{x}^1 = (1, 2)$ with $(2, 1)$, condition (9) requires $0.5\bar{u}(1) + 0.5\bar{u}(2)$ to be strictly greater than itself, which is impossible.²⁸

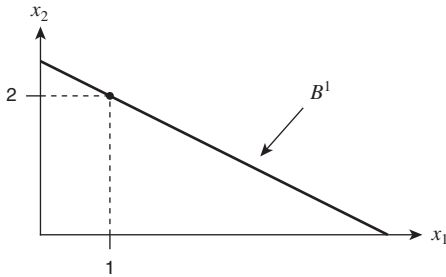
Afriat's Theorem characterizes classical datasets that are rationalizable by locally nonsatiated utility functions. Readers who are familiar with Afriat's Theorem will notice some similarity between it and Theorem 1, in the sense that both results involve revealed preference relationships (such as (6) and (7), or the EU versions (8) and (9)), between the chosen bundle \mathbf{x}^t and a finite subset of the budget set \underline{B}^t . In the case of Theorem 1 this subset is $\underline{B}^t \cap \mathcal{G}$ whereas in the case of Afriat's Theorem, the comparison is with $\underline{B}^t \cap \mathcal{D}$, where $\mathcal{D} = \{\mathbf{x}^t\}_{t=1}^T$. Theorem 1 requires a relaxation of the domain restriction used in Afriat's Theorem (\mathcal{G} , which contains \mathcal{D}) because it characterizes rationalizability by utility functions with added structure. This is clear

²⁶ See the remarks at the end of Section IA.

²⁷ If $\pi_s^t = 1/2$ for all t, s , the reader can verify that one solution to this problem is $\bar{u}(0) = 0$, $\bar{u}(1) = 1$, $\bar{u}(2) = 4$, $\bar{u}(3) = 6$, $\bar{u}(4) = 8$, and $\bar{u}(6) = 9$, i.e., the dataset depicted in panel A of Figure 1 is EU-rationalizable.

²⁸ If the state probabilities are not known to the observer then it is impossible to reject expected utility with only one observation. Instead the observation in Example 1 would tell us that state 2 is more probable than state 1. This means that if there is another observation where the subject buys more of state 1 consumption even though it is more expensive, an observer could conclude that the agent is *not* maximizing expected utility. This is the essential idea in Epstein (2000).

Panel A. Observation 1



Panel B. Observation 1 with grid inserted

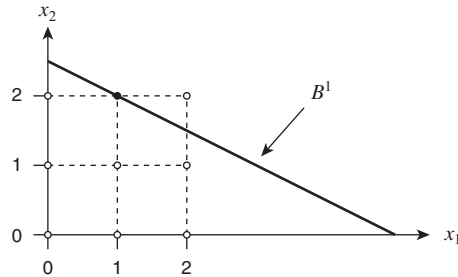


FIGURE 2. THE GRID METHOD APPLIED TO EXAMPLE 1

from Example 1, where the observation is *not* EU-rationalizable, even though the inequalities (8) and (9) are trivially satisfied in $\underline{B}^1 \cap \mathcal{D}$, since $\mathcal{D} = \{\mathbf{x}^1\}$.²⁹

There is a further connection between the two results. In the case of Afriat's Theorem, the revealed preference relations can be formulated as a no-cycling condition among the elements of \mathcal{D} called the Generalized Axiom of Revealed Preference (GARP). By appealing to a result of Fishburn (1975), it is possible to characterize EU-rationalizability in terms of a condition that is stronger than GARP but similar to it in the sense that it forbids a generalized notion of a revealed preference cycle on the set \mathcal{G} . These observations are discussed in greater detail in online Appendix Section A1.2.

C. Making Out-Of-Sample Predictions

Once it has been ascertained that a subject's behavior is consistent with a given model, it would be natural to exploit this compatibility by using the same model to make predictions of that subject's out-of-sample behavior. We now explain how this can also be done using the GRID method. To simplify our discussion, we only explain this in the context of the EU model; making predictions when some other model is assumed can be carried out in a similar fashion. The procedure outlined here is implemented in Section IVB.

Suppose $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is collected from a subject who is EU-rationalizable (with objective probabilities $\pi_s^t > 0$ for all t and s). Using the information from \mathcal{O} and assuming that the subject is behaving as an EU-maximizer, how can we predict the subject's choice between two lotteries: lottery A, which pays out a_i with probability $\alpha_i > 0$ (for $i = 1, 2$, with $\alpha_1 + \alpha_2 = 1$), and lottery B which pays out b_i with probability $\beta_i > 0$ (for $i = 1, 2$, with $\beta_1 + \beta_2 = 1$)? In formal terms, a strict preference for A over B is consistent with the EU model if there is a Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ which EU-rationalizes \mathcal{O} and satisfies $\alpha_1 u(a_1) + \alpha_2 u(a_2) > \beta_1 u(b_1) + \beta_2 u(b_2)$. Whether u exists can easily be answered using Theorem 1. In this case, it is convenient to choose \mathcal{X} to be strictly larger than \mathcal{X}^* . Specifically, let $\mathcal{X} = \mathcal{X}^* \cup \{a_1, a_2, b_1, b_2\}$. Since \mathcal{O} is EU-rationalizable, there must be a

²⁹For other results which involve comparing \mathbf{x}^t with a subset of \underline{B}^t that is larger than $\underline{B}^t \cap \mathcal{D}$, see footnote 23.

strictly increasing function $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ that solves the inequalities (8) and (9). Furthermore, Theorem 1 tells us that \bar{u} has an extension u , with domain \mathbb{R}_+ , that rationalizes \mathcal{O} . Therefore, to ascertain whether a strict preference for A over B is consistent with the EU model, a necessary and sufficient test is whether there is a strictly increasing function $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ that, in addition to (8) and (9), obeys

$$(10) \quad \alpha_1 \bar{u}(a_1) + \alpha_2 \bar{u}(a_2) > \beta_1 \bar{u}(b_1) + \beta_2 \bar{u}(b_2).$$

This test is easy to implement since (10) is a linear inequality. Note that because there are potentially multiple Bernoulli functions that EU-rationalize the data, it is entirely possible that *both* a strict preference for A over B and a strict preference for B over A are consistent with the EU model: in this case, there will be an increasing function \bar{u} that solves (8), (9), and (10), and another one that solves (8), (9), and (10), the last with the inequality reversed.

D. Testing Other Models Using the GRID Method

So far, we have considered tests of EU-rationalizability in the case where the probability of each state is known to both the agent and the observer. Our test could be extended to the case where no objective probabilities can be attached to each state. A dataset $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ is *rationalizable by subjective expected utility* (SEU) if there exists probability distribution $\pi = (\pi_1, \pi_2, \dots, \pi_{\bar{s}}) \gg 0$ and a Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that, at every observation t , we have $\sum_{s=1}^{\bar{s}} \pi_s u(x_s^t) \geq \sum_{s=1}^{\bar{s}} \pi_s u(x_s)$ for all $\mathbf{x} \in B^t$. In this case, ϕ is independent of t and is required to belong to the family Φ_{SEU} such that $\phi \in \Phi_{SEU}$ if $\phi(\mathbf{u}) = \sum_{s=1}^{\bar{s}} \pi_s u_s$ for some $\pi \gg 0$. By Theorem 1, $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ can be rationalized by some $\phi \in \Phi_{SEU}$ if and only if *there is a strictly increasing \bar{u} such that (8) and (9) hold for some $\pi \gg 0$* . These conditions form a system of inequalities bilinear in the unknowns $\{\pi_s\}_{s=1}^{\bar{s}}$ and $\{\bar{u}(r)\}_{r \in \mathcal{X}}$.

For many of the standard models of decision making under risk, under uncertainty, or over time, the rationalizability problem has a structure similar to that of SEU in the sense that it involves finding a Bernoulli function u and an aggregator function ϕ belonging to some family Φ that together rationalize the data, and this problem can in turn be transformed via Theorem 1 into a problem of solving a system of bilinear inequalities. In online Appendix Section A3, we use Theorem 1 to devise such tests for various models of contingent choice, including choice acclimating personal equilibrium (Kőszegi and Rabin 2007), maxmin expected utility (Gilboa and Schmeidler 1989), variational preferences (Maccheroni, Marinacci, and Rustichini 2006), and a model with budget-dependent reference points. We also explain how we could test models of choice over time on data from budgetary allocations, such as those collected by Andreoni and Sprenger (2012). A model of discounted utility (with or without present bias) is formally very similar to the subjective expected utility model.

Even though solving a bilinear problem may be computationally intensive, the Tarski-Seidenberg Theorem tells us that this problem is decidable, in the sense that there is a known algorithm that can determine in a finite number of steps whether a solution exists. Nonlinear tests are not new to the revealed preference literature;

for example, they appear in tests of weak separability (Varian 1983a), in tests of maxmin expected utility and other models of ambiguity (Bayer et al. 2013), and in tests of Walrasian general equilibrium (Brown and Matzkin 1996). Solving these problems can be computationally straightforward in some cases because of special features of the model/environment or when the number of observations is small. The tests that we develop simplify dramatically and are easily implementable in practice when there are only two states (though they remain nonlinear).

The two-state case, while special, is very common in applied theoretical settings and laboratory experiments. For example, to implement the SEU test, we simply condition on the probability of state 1 (and hence on the probability of state 2), and then perform a linear test to check whether there is a strictly increasing function \bar{u} solving (8) and (9). If not, we choose another probability for state 1, implement the test, and repeat (if necessary). Even a uniform grid search of up to two decimal places on the probability of state 1 will lead to no more than 99 linear tests, which can be implemented with very little difficulty.³⁰

Rank-Dependent Utility (RDU).—In Section IV, we report the findings of an empirical test of the RDU model (Quiggin 1982) when there are two states, so we explain this case of the model, and its corresponding test, in detail here. The online Appendix contains a full treatment of the multi-state case (see Sections A2 and A4).

Consider an experiment where the probabilities of states 1 and 2 are objectively known and given by $\pi_1 > 0$ and $\pi_2 > 0$. With no loss of generality, assume that $\pi_1 \geq \pi_2$. In the RDU model, the subject behaves as though these probabilities are distorted: if state s is the less favorable state, i.e., the state where the payoff is smaller, then the weight given to state s is ρ_s , with this distortion respecting the rank of the objective probabilities, i.e.,

$$(11) \quad 1 > \rho_1 > \rho_2 > 0 \quad \text{if } \pi_1 > \pi_2 \quad \text{and} \quad 1 > \rho_1 = \rho_2 > 0 \quad \text{if } \pi_1 = \pi_2.$$

The utility of (x_1, x_2) when $x_1 \leq x_2$ is $V(x_1, x_2) = \rho_1 u(x_1) + (1 - \rho_1)u(x_2)$ and the utility of (x_1, x_2) when $x_1 > x_2$ is $V(x_1, x_2) = (1 - \rho_2)u(x_1) + \rho_2 u(x_2)$. Converting this into the framework of Theorem 1, we are testing rationalizability in the case where ϕ has the form

$$(12) \quad \phi(u_1, u_2) = \begin{cases} \rho_1 u_1 + (1 - \rho_1)u_2 & \text{if } u_1 \leq u_2 \\ (1 - \rho_2)u_1 + \rho_2 u_2 & \text{if } u_1 > u_2 \end{cases}.$$

By Theorem 1, a necessary and sufficient condition for $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ to be RDU-rationalizable is for there to be ρ_1 and ρ_2 obeying (11) and a strictly increasing \bar{u} such that with ϕ defined by (12), the conditions (6) and (7) admits a solution. Given the formula for ϕ , this test involves solving a set of inequalities that are bilinear in the unknowns $\{\bar{u}(r)\}_{r \in \mathcal{X}}$ and $\{\rho_1, \rho_2\}$. When implementing this test, we let ρ_1 and ρ_2 take different values on a very fine grid in $[0, 1]^2$, subject to (11), and (for each

³⁰While we have not found it necessary to use them in our implementation in this paper, there are solvers available for mixed integer nonlinear programs (for example, as surveyed in Bussieck and Vigerske 2010) that are potentially useful for implementing bilinear tests more generally.

TABLE 2—CONDITIONS ON \bar{u} FOR RDU-RATIONALIZABILITY GIVEN (ρ_1, ρ_2)

| Monotonicity of \bar{u} | Optimality of $\mathbf{x}^1 = (2, 4)$ |
|--|---|
| $\bar{u}(6) > \bar{u}(4) > \bar{u}(3) > \bar{u}(2)$ $\bar{u}(2) > \bar{u}(1) > \bar{u}(0)$ | $\rho_1 \bar{u}(2) + (1 - \rho_1) \bar{u}(4) > \rho_1 \bar{u}(1) + (1 - \rho_1) \bar{u}(6)$ $\rho_1 \bar{u}(2) + (1 - \rho_1) \bar{u}(4) \geq (1 - \rho_2) \bar{u}(3) + \rho_2 \bar{u}(1)$ |
| Optimality of $\mathbf{x}^2 = (6, 1)$ | Optimality of $\mathbf{x}^3 = (4, 3)$ |
| $(1 - \rho_2) \bar{u}(6) + \rho_2 \bar{u}(1) \geq \rho_1 \bar{u}(0) + (1 - \rho_1) \bar{u}(3)$ $(1 - \rho_2) \bar{u}(6) + \rho_2 \bar{u}(1) \geq (1 - \rho_2) \bar{u}(3) + \rho_2 \bar{u}(2)$ | $(1 - \rho_2) \bar{u}(4) + \rho_2 \bar{u}(3) > \rho_1 \bar{u}(2) + (1 - \rho_1) \bar{u}(4)$ $(1 - \rho_2) \bar{u}(4) + \rho_2 \bar{u}(3) > (1 - \rho_2) \bar{u}(6) + \rho_2 \bar{u}(1)$ |

case) perform the corresponding linear test to search for a solution in $\{\bar{u}(r)\}_{r \in \mathcal{X}}$; \mathcal{O} is RDU-rationalizable if such a solution exists for some values of ρ_1 and ρ_2 .

As an illustration, consider again the dataset displayed in panel A of Figure 1. In Table 2 we collect the relevant inequalities for rationalizability by ϕ as defined by (12); the dataset can be rationalized by ϕ (for specific values of ρ_1 and ρ_2) if and only if there is \bar{u} that satisfies the inequalities displayed in the table. Comparing this test with the test for EU-rationalizability (displayed in Table 1), notice that there is no change to \mathcal{X} or to \mathcal{G} , nor is there a change to the relevant comparisons at each observation (for example at observation 1, $(2, 4)$ is compared against $(1, 6)$ and $(3, 1)$ in both tables). The difference between them is in the functional form, with the EU-form in Table 1 and the RDU-form in Table 2.

Disappointment Aversion (DA).—We also implement a GRID test of the DA model (Gul 1991). When there are two states, the DA model is a special case of RDU, with a further restriction on ρ_1 and ρ_2 . Specifically, there is $\beta \in (-1, \infty)$ such that, for $s = 1, 2$,

$$(13) \quad \rho_s = \frac{(1 + \beta) \pi_s}{1 + \pi_s \beta}.$$

Note that this restriction has bite only if $\pi_1 \neq \pi_2$, so the RDU and DA models coincide when $\pi_1 = \pi_2$. If $\beta = 0$, the agent simply maximizes expected utility. If $\beta > 0$, we have $\rho_s > \pi_s$; the agent attaches a probability to state s that is higher than the objective probability when state s is the less favorable state and the agent is said to be *disappointment averse*. If $\beta < 0$, then $\rho_s < \pi_s$, and the agent is *elation seeking*. Similar to RDU, we test the DA model by letting β take on different values and performing the corresponding linear test.³¹

While it is well known that the RDU and EU models lead to different predictions, it is not immediately clear that they are observationally distinct in the context of observations drawn from linear budgets. We end this section with an example of a dataset that is RDU-rationalizable but not EU-rationalizable.

Example 2: Suppose the dataset consists of three observations as depicted in Figure 3 where $\mathbf{x}^1 = (a, a)$, $\mathbf{x}^2 = (b, b)$, and $\mathbf{x}^3 = (c, d)$. Note that (b, c) is on the first observation's budget line and (a, d) is on the second observation's budget line.

³¹ In practice, we let ρ_1 take on different values on $(0, 1)$, back out the corresponding value of β (according to (13)) and then work out ρ_2 (with (13)).

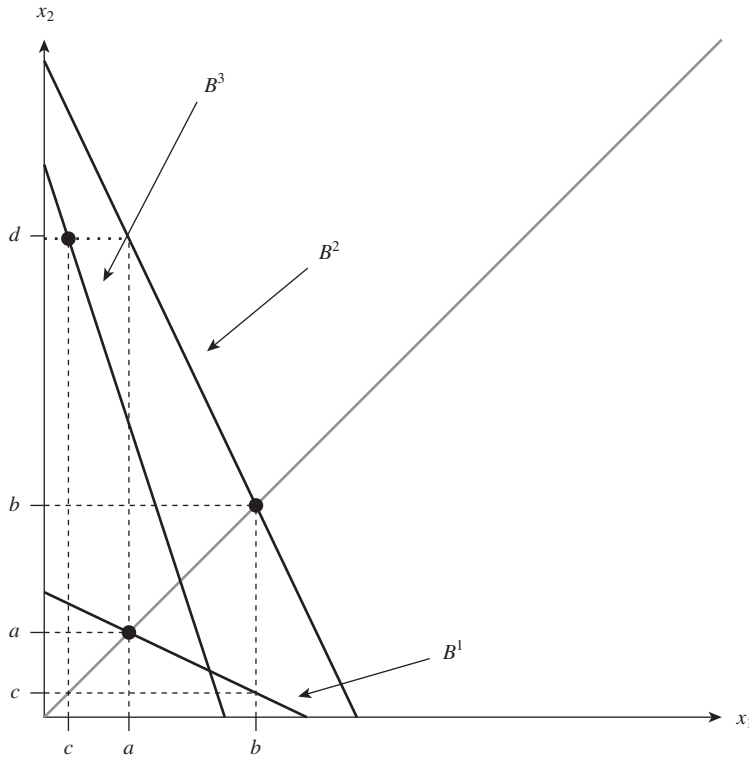


FIGURE 3. RDU-RATIONALIZABLE BUT NOT EU-RATIONALIZABLE DATASET

The price at observation t is $\mathbf{p}^t = (1, q^t)$, where $q^1 > 1 > 1/q^1 \geq q^2 > q^3$. Consequently the first budget line is the flattest and the third budget line is the steepest.

We claim that these observations are not EU-rationalizable if the two states are equiprobable. Suppose that they are, for some Bernoulli function u . Then the first observation tells us that $2u(a) \geq u(b) + u(c)$, since (b, c) is available when (a, a) is chosen. Similarly, from the second observation, we know that $2u(b) \geq u(a) + u(d)$. Together this gives

$$u(b) - u(d) \geq u(a) - u(b) \geq u(c) - u(a),$$

from which we obtain $u(a) + u(b) \geq u(c) + u(d)$. But this is contradicted by observation 3 where (c, d) is chosen even though (a, b) is in the interior of the budget set.

However, these observations are RDU-rationalizable. This should be quite intuitive because the demand pattern involves stickiness on the 45-degree line over a range of prices, with the demand deviating away from $x_1 = x_2$ only when (at the third observation) state 2 consumption is sufficiently cheap. Indeed, suppose $V(x_1, x_2) = \rho u(x_1) + (1 - \rho)u(x_2)$ when $x_1 \leq x_2$ and $V(x_1, x_2) = (1 - \rho)u(x_1) + \rho u(x_2)$ when $x_1 > x_2$, with $\rho \in (1/2, 1)$, so the agent displays disappointment aversion. It is straightforward to check that if u is strictly concave, then the agent's utility is maximized at $x_1 = x_2$ at observations 1 and 2 so long as $\rho \geq 1/(1 + q^2)$ and $\rho \geq q^1/(1 + q^1)$. Since we assume that $1/q^1 \geq q^2$, the

first inequality is tighter than the second. Let us set $\rho = 1/(1 + q^2)$. It remains for us to find a Bernoulli function that rationalizes the third observation. It suffices to find $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $u' > 0$ and $u'' < 0$ and $u'(c)$ and $u'(d)$ satisfy the first-order condition

$$\frac{\rho u'(c)}{(1 - \rho)u'(d)} = \frac{1}{q^2} \frac{u'(c)}{u'(d)} = \frac{1}{q^3}.$$

Since $d > c$ and $q^2 > q^3$, such a function u must exist.

II. Measuring Departures from Rationalizability

The revealed preference tests presented in the previous section are “sharp,” in the sense that a dataset either passes the test for a given model or it fails. This either/or feature of the tests is not particular to our results but is true of all classical revealed preference tests, including Afriat’s. It would, of course, be desirable to develop a way of measuring *the extent* to which a given class of utility functions succeeds or fails in rationalizing a dataset, and the most common approach adopted in the revealed preference literature to address this issue was developed by Afriat (1972, 1973) and Varian (1990) in the context of classical datasets, i.e., datasets with classical budget sets (see (1)).^{32,33} The basic idea is that if a consumer’s choice fails to maximize utility, then it is natural to compare what he spent with what he *could* have spent in order to achieve the same utility level. This gives us a metric to quantify the utility loss in expenditure terms. We now give an account of this approach and explain how to use the GRID method to calculate this index.

Let $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ be a classical dataset. For any number $e^t \in [0, 1]$, we define

$$(14) \quad B^t(e^t) = \mathcal{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t) \cup \{\mathbf{x}^t\} = \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p}^t \cdot \mathbf{x} \leq e^t \mathbf{p}^t \cdot \mathbf{x}^t\} \cup \{\mathbf{x}^t\}.$$

Notice that when $e^t = 1$, this set coincides with the true budget set $\mathcal{B}(\mathbf{p}^t, \mathbf{p}^t \cdot \mathbf{x}^t)$ (see (1)). If $e^t < 1$, then $B^t(e^t)$ is a shrunken version of this set that retains the observed choice \mathbf{x}^t but removes all bundles for which total expenditure at \mathbf{p}^t is strictly higher than $e^t \mathbf{p}^t \cdot \mathbf{x}^t$, i.e., those bundles \mathbf{x} where $\mathbf{p}^t \cdot \mathbf{x} > e^t \mathbf{p}^t \cdot \mathbf{x}^t$. Clearly $B^t(e^t)$ shrinks with the value of e^t . Given $\mathbf{e} = (e^t)_{t=1}^T$, we refer to $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t))\}_{t=1}^T$ as a *modified dataset*.

Let \mathcal{U} be a collection of utility functions defined on $\mathbb{R}_+^{\bar{s}}$ belonging to a given family; for example, \mathcal{U} could be the family of locally nonsatiated utility functions, which was the family considered by Afriat (1972, 1973) and Varian (1990). The modified dataset $\mathcal{O}(\mathbf{e})$ is *rationalizable by \mathcal{U}* if there is $U \in \mathcal{U}$ such that $U(\mathbf{x}^t) \geq U(\mathbf{x})$ for all $\mathbf{x} \in B^t(e^t)$. Clearly, if $\mathcal{O}(\mathbf{e})$ is rationalizable by \mathcal{U} , then so is

³²For examples where Afriat-Varian type indices are used to measure a model’s fit, see Mattei (2000); Harbaugh, Krause, and Berry (2001); Andreoni and Miller (2002); Choi et al. (2007, 2014); Beatty and Crawford (2011); and Halevy, Persitz, and Zrill (2018). See also Echenique, Lee, and Shum (2011), which develops and applies a related index called the money pump index.

³³Varian (1990) and Halevy, Persitz, and Zrill (2018) discuss why such measures may be more suitable than other measures such as the sum of squared errors between observed and predicted demands.

$\mathcal{O}(\mathbf{e}')$ for any $\mathbf{e}' < \mathbf{e}$, since shrinking budget sets will make it easier for rationalizability to hold. Furthermore, if $U(\mathbf{x}) \geq U(\mathbf{0})$ for all $\mathbf{x} \geq \mathbf{0}$ for some $U \in \mathcal{U}$, then $\mathcal{O}(\mathbf{e})$ is rationalizable at $\mathbf{e} = \mathbf{0}$ (since $B^t(\mathbf{0}) = \{\mathbf{0}, \mathbf{x}^t\}$ for all t), though in general $\mathcal{O}(\mathbf{e})$ will be rationalizable by \mathcal{U} without shrinking budget sets so drastically. This suggests that if $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ is not rationalizable by \mathcal{U} , then one way of measuring the severity of this failure is to measure the extent to which budget sets need to shrink to obtain rationalizability. This is the key idea behind the indices proposed by Afriat and Varian.

Afriat's proposal is to shrink all budget sets by the *same* factor e (so $\mathbf{e} = (e, e, \dots, e)$) and to find the largest number e at which $\mathcal{O}(\mathbf{e})$ is rationalizable by \mathcal{U} . Afriat refers to

$$\sup \{e : \mathcal{O}(e, e, \dots, e) \text{ is rationalizable by } \mathcal{U}\}$$

as the *critical cost efficiency index*; we shall also refer to it as Afriat's efficiency index or Afriat's index. Of course if $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ is itself rationalizable by \mathcal{U} , then this index equals 1. If this index equals $e^* < 1$, then it means that there is some utility function in \mathcal{U} for which the observed choice \mathbf{x}^t is superior to every bundle that costs less than $e^* \mathbf{p}^t \cdot \mathbf{x}^t$, but rationality is limited because there is some observation t' and a bundle \mathbf{y} costing more than $e^* \mathbf{p}^{t'} \cdot \mathbf{x}^{t'}$ but less than $\mathbf{p}^{t'} \cdot \mathbf{x}^{t'}$ that gives strictly higher utility than $\mathbf{x}^{t'}$.

The alternative measure proposed by Varian (1990) allows different budget sets to shrink by different factors; the degree to which a dataset is not consistent with a particular model is then given by the smallest sum of square differences between the *efficiency vector* \mathbf{e} and the vector $(1, 1, \dots, 1)$, i.e., $\sum_{t=1}^T (1 - e^t)^2$. Renormalizing this measure to facilitate comparison with Afriat's index, we define *Varian's efficiency index* as

$$\sup \left\{ 1 - \sqrt{\frac{\sum_{t=1}^T (1 - e^t)^2}{T}} : \mathcal{O}(\mathbf{e}) \text{ is rationalizable by } \mathcal{U} \right\}.$$

When \mathcal{O} is rationalizable by \mathcal{U} , both Varian's index and Afriat's equal 1. In both cases, a higher number indicates greater consistency with a model, but Varian's index is always weakly higher than Afriat's since it maximizes over a greater set of efficiency vectors.

Even though Varian's index is in a sense more discriminating than the one proposed by Afriat, Afriat's measure is more commonly used because it is much easier to compute: while calculating the latter simply requires searching for a threshold e^* at which the modified dataset is just rationalizable by \mathcal{U} ,³⁴ calculating Varian's index requires searching through all efficiency vectors \mathbf{e} .³⁵ In our empirical analysis in Section IV, we use Afriat's efficiency index, because it is easy to compute and

³⁴The binary search algorithm works as follows. We first set the lower and upper bounds on e^* to $e_L = 0$ and $e_H = 1$, respectively. We then check whether the dataset passes or fails the test at $e = (e_L + e_H)/2$; if it passes the test, then we update both e^* and its lower bound to $(e_L + e_H)/2$; if it fails the test, then we update e^* to e_L and the upper bound on e^* to $(e_L + e_H)/2$. We then repeat the procedure, selecting and testing the new midpoint of the updated lower and upper bounds. The algorithm terminates when the lower and upper bounds are sufficiently close, in our case within 10^{-6} of one another.

³⁵In fact, it is known that calculating Varian's index is an NP hard problem (Smeulders et al. 2014).

because it facilitates comparison with other papers, which mostly use this index. But we also calculate Varian's index for some (though not all) of the models we consider, using a new algorithm we developed that searches for the optimal efficiency vector \mathbf{e} (see online Appendix Sections A8 and A9).

A. Testing the EU-Rationalizability of $\mathcal{O}(\mathbf{e})$

Whether one is calculating Afriat's index or Varian's, it will require checking if a modified dataset $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t))\}_{t=1}^T$ is rationalizable by \mathcal{U} . When \mathcal{U} is the family of all locally nonsatiated utility functions, a generalization of Afriat's Theorem provides a test for the rationalizability of $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t))\}_{t=1}^T$ (described in detail in online Appendix Section A5.1). In the case where \mathcal{U} is the family of expected utility or rank-dependent utility functions, we can perform a GRID test.

To be specific, consider a dataset $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$ collected from an experiment in which state s occurs with probability $\pi_s > 0$ at every observation. Suppose that for a given $\mathbf{e} = (e^t)_{t=1}^T$, we would like to check whether $\mathcal{O}(\mathbf{e})$ is EU-rationalizable or, in the language of this section, whether it is rationalizable by \mathcal{U} , where $U : \mathbb{R}_+^{\bar{s}} \rightarrow \mathbb{R}_+$ is in \mathcal{U} if $U(\mathbf{x}) = \sum_{s=1}^{\bar{s}} \pi_s u(x_s)$ for some Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. By Theorem 1, a necessary and sufficient condition for rationalizability is that there is a strictly increasing function $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ that satisfies (8) and (9), which form a set of linear conditions. Note that the constraint set at observation t is $B^t(e^t)$ (which is *not* a linear or even convex budget set) and its downward extension is $\{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{p}^t \cdot \mathbf{x} \leq e^t \mathbf{p}^t \cdot \mathbf{x}^t\} \cup \{\mathbf{x} \in \mathbb{R}_+^{\bar{s}} : \mathbf{x} \leq \mathbf{x}^t\}$.

As an illustration, we return to the example first depicted in panel A of Figure 1 and suppose we shrink B^1 and B^2 by $e^1, e^2 < 1$ respectively but leave B^3 as it is (so that $\mathbf{e} = (e^1, e^2, 1)$), as shown in panel A of Figure 4. The downward extensions of $B^1(e^1)$, $B^2(e^2)$, are depicted in panel B, along with B^3 (which is unchanged and coincides with its downward extension). In this case, $\mathcal{X}^* = \{0, 1, 2, 3, 4, 6\}$, we can choose $\mathcal{X} = \mathcal{X}^*$, and (by Theorem 1) $\mathcal{O}(\mathbf{e})$ is EU-rationalizable if and only if there is \bar{u} that solves the inequalities in Table 3. We again have the strict monotonicity conditions in the top-left panel, with the other panels listing the optimality conditions applicable to \mathbf{x}^1 , \mathbf{x}^2 , and \mathbf{x}^3 . There are 17 bundles in $\underline{B}^1(e^1) \cap \mathcal{G}$ (besides the observed choice $\mathbf{x}^1 = (2, 4)$); two of them, $(1, 6)$ and $(3, 0)$, are undominated and both lie on the upper boundary, which leads (by condition (8)) to the weak inequalities displayed in the top-right panel. There is just one undominated bundle, $(2, 2)$, in $\underline{B}^2(e^2) \cap \mathcal{G}$ (besides \mathbf{x}^2); this leads to the strict inequality displayed in the bottom-left panel (by condition (9)). The optimality conditions on \mathbf{x}^3 are essentially unchanged from those displayed in Table 1.

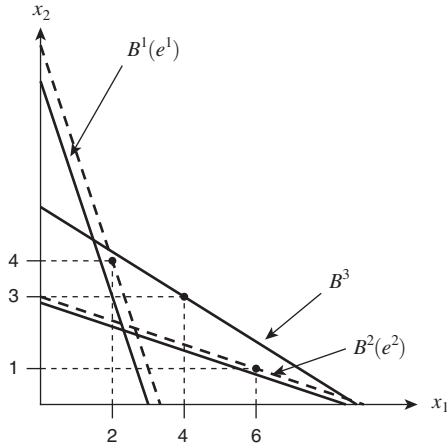
Checking for the RDU-rationalizability of $\mathcal{O}(\mathbf{e})$ involves a similar procedure, with the functional form modified as explained in Section ID.

B. Approximate Smooth Rationalizability

While Theorem 1 guarantees that there is a Bernoulli function u that extends $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ and rationalizes the data when the required conditions are satisfied, the Bernoulli function is not necessarily smooth. The smoothness of u is commonly

TABLE 3—CONDITIONS ON \bar{u} FOR THE EU-RATIONALIZABILITY OF $\mathcal{O}(\mathbf{e})$

| Monotonicity of \bar{u} | Optimality of $\mathbf{x}^1 = (2, 4)$ |
|---|--|
| $\bar{u}(6) > \bar{u}(4) > \bar{u}(3) > \bar{u}(2)$ | $\pi_1 \bar{u}(2) + \pi_2 \bar{u}(4) \geq \pi_1 \bar{u}(1) + \pi_2 \bar{u}(6)$ |
| $\bar{u}(2) > \bar{u}(1) > \bar{u}(0)$ | $\pi_1 \bar{u}(2) + \pi_2 \bar{u}(4) \geq \pi_1 \bar{u}(3) + \pi_2 \bar{u}(0)$ |
| Optimality of $\mathbf{x}^2 = (6, 1)$ | Optimality of $\mathbf{x}^3 = (4, 3)$ |
| $\pi_1 \bar{u}(6) + \pi_2 \bar{u}(1) > \pi_1 \bar{u}(2) + \pi_2 \bar{u}(2)$ | $\pi_1 \bar{u}(4) + \pi_2 \bar{u}(3) > \pi_1 \bar{u}(2) + \pi_2 \bar{u}(4)$ |
| | $\pi_1 \bar{u}(4) + \pi_2 \bar{u}(3) > \pi_1 \bar{u}(6) + \pi_2 \bar{u}(1)$ |

Panel A. Modified dataset with $\mathbf{e} = (e^1, e^2, 1)$ 

Panel B. Modified dataset with grid inserted

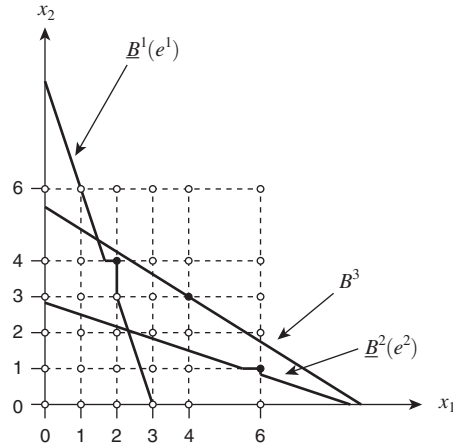


FIGURE 4. THE GRID METHOD APPLIED TO A MODIFIED DATASET

assumed in applications of expected utility and related models and its implications can appear to be stark. For example, suppose that it is commonly known that states 1 and 2 occur with equal probability and that we observe the agent choosing $(1, 1)$ at a price vector (p_1, p_2) , with $p_1 \neq p_2$. This observation is incompatible with a smooth EU model; indeed, given that the two states are equiprobable, the slope of the indifference curve at $(1, 1)$ must equal -1 and thus it will not be tangential to the budget line and will not be a local optimum. On the other hand, it is trivial to check that this observation is EU-rationalizable in our sense. In fact, one could even find a *concave* Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ for which $(1, 1)$ maximizes expected utility. (Such a u will be continuous and strictly increasing, but have a kink at 1.)

These two facts are reconcilable. Given any strictly increasing and continuous function u defined on a compact interval of \mathbb{R}_+ , there is a strictly increasing and smooth function \bar{u} that is uniformly and arbitrarily close to u on that interval. Thus, if a Bernoulli function u rationalizes $\mathcal{O} = \{(\mathbf{x}^t, B^t)\}_{t=1}^T$ by $\{\phi(\cdot, t)\}_{t=1}^T$, then for any efficiency threshold $e \in (0, 1)$, there is a *smooth* Bernoulli function \bar{u} that rationalizes $\mathcal{O}' = \{(\mathbf{x}^t, B^t(e))\}_{t=1}^T$ by $\{\phi(\cdot, t)\}_{t=1}^T$. In other words, if a dataset is rationalizable by a Bernoulli function, then it can also be rationalized by a smooth Bernoulli function, for any efficiency threshold arbitrarily close to 1. In this sense, imposing a smoothness requirement on the Bernoulli function does not radically alter a model's ability to explain a dataset.

III. Concave Bernoulli Functions

A common assumption in applications of expected utility (EU) theory is that agents are risk averse, which is equivalent to the concavity of the Bernoulli function. The necessary and sufficient conditions that we have developed for EU-rationalizability (in Theorem 1) neither require nor guarantee that the Bernoulli function is concave. This distinction is significant because there are datasets which can be rationalized by the EU model, but only with nonconcave Bernoulli functions. This will be made readily apparent in the empirical implementation in Section IV, but we also provide an intuitive example of such a phenomenon, in a classical dataset with two observations, in online Appendix Section A4.1.

In this section we provide a test for *concave EU-rationalizability*, i.e., EU-rationalizability with a concave Bernoulli function. (Recall that, by definition, the Bernoulli function is continuous and strictly increasing.) Unfortunately, we do not, in this case, have a result like Theorem 1 which is applicable to observations drawn from general compact constraint sets. Our procedure works in a narrower set of environments: it allows us to test for the concave EU-rationalizability of a classical dataset $\mathcal{O} = \{(\mathbf{x}^t, \mathbf{p}^t)\}_{t=1}^T$, and also when it is modified by $\mathbf{e} = (e^t)_{t=1}^T$, i.e., the dataset $\mathcal{O}(\mathbf{e}) = \{(\mathbf{x}^t, B^t(e^t))\}_{t=1}^T$ (with $B^t(e^t)$ defined by (14)). Note that there is already a test of concave EU-rationalizability for classical datasets (see Varian 1983a and Green and Srivastava 1986), but that test makes use of the *sufficiency* of the first-order conditions, which in turn relies crucially on the linearity of the classical budget sets; since $B^t(e^t)$ is *not* a convex set, that method does not obviously extend to testing for the concave EU-rationalizability of $\mathcal{O}(\mathbf{e})$. The added value of our approach lies in its applicability to modified datasets, which (as we explain in Section II) enables us to calculate the critical cost efficiency index when \mathcal{O} itself is not concave EU-rationalizable.

Our test for concave EU-rationalizability is another instance of a GRID test: we identify, in the modified budget set $B^t(e^t)$, a *finite* number of bundles for which the superiority of the chosen bundle \mathbf{x}^t over these bundles is *sufficient* to guarantee the optimality of \mathbf{x}^t over all bundles in $B^t(e^t)$. However, the set of comparison bundles will be chosen differently from the case where concavity is not required (it is no longer $\mathcal{G} \cap B^t(e^t)$).

We shall confine our discussion in this section to the test for concave EU-rationalizability when there are two states with commonly known, strictly positive probabilities π_1 and π_2 ; the test when there are multiple states is covered in online Appendix Section A4.3. (Tests for the RDU model with a concave Bernoulli function are presented in Sections A4.2 and A4.3.) To explain this test we need some definitions and a key observation.

Let $\bar{r} > 0$ be a real number such that $\bar{r}p_s^t \geq p^t \cdot x^t$ for all t and s . Notice that, in any state and at any observation, the subject cannot afford to choose a consumption level strictly higher than \bar{r} . Thus, for the purposes of rationalizing a dataset, the behavior of the Bernoulli function beyond \bar{r} is of no relevance. We define

$$(15) \quad \mathcal{X}^{**} = \{0\} \cup \{r \in \mathbb{R}_+ : r = x_s^t \text{ for some } t \text{ and } s\} \cup \{\bar{r}\}$$

and let $\mathcal{X} \subset \mathbb{R}_+$ be a finite set containing \mathcal{X}^{**} .³⁶

For any function $\bar{h} : \mathcal{X} \rightarrow \mathbb{R}_+$, we define the *piecewise linear extension* of \bar{h} (or simply *linear extension* of \bar{h} for short) as the function $\bar{h}_\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that is linear between adjacent points in \mathcal{X} , with $\bar{h}_\ell(r) = \bar{h}(r)$ for all $r \in \mathcal{X}$.^{37,38} Thus, if $r < r'$ are adjacent points in \mathcal{X} ,

$$\bar{h}_\ell(a) = \lambda \bar{h}(r) + (1 - \lambda) \bar{h}(r') \quad \text{whenever } a = \lambda r + (1 - \lambda) r' \text{ for } \lambda \in (0, 1).$$

The importance of piecewise linear functions comes from the following key observation:

Suppose $\mathcal{O}(\mathbf{e})$ is EU-rationalizable by the concave Bernoulli function u and let $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ be the restriction of u to \mathcal{X} . Then $\mathcal{O}(\mathbf{e})$ is EU-rationalizable by $\bar{u}_\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, the linear extension of \bar{u} .

To see why this is true, note that $\bar{u}_\ell(r) = u(r)$ for all $r \in \mathcal{X}$ and the concavity of u guarantees that $u(r) \geq \bar{u}_\ell(r)$ for all $r \in [0, \bar{r}]$. Thus, $\pi_1 u(x_1^t) + \pi_2 u(x_2^t) = \pi_1 \bar{u}_\ell(x_1^t) + \pi_2 \bar{u}_\ell(x_2^t)$ for all $\mathbf{x}^t = (x_1^t, x_2^t)$ and, for any other bundle $\mathbf{x} = (x_1, x_2)$, we have $\pi_1 u(x_1) + \pi_2 u(x_2) \geq \pi_1 \bar{u}_\ell(x_1) + \pi_2 \bar{u}_\ell(x_2)$. Since

$$\pi_1 u(x_1^t) + \pi_2 u(x_2^t) \geq \pi_1 u(x_1) + \pi_2 u(x_2) \quad \text{for all } \mathbf{x} = (x_1, x_2) \in B^t(e^t),$$

we also have

$$\pi_1 \bar{u}_\ell(x_1^t) + \pi_2 \bar{u}_\ell(x_2^t) \geq \pi_1 \bar{u}_\ell(x_1) + \pi_2 \bar{u}_\ell(x_2) \quad \text{for all } \mathbf{x} = (x_1, x_2) \in B^t(e^t),$$

which completes the proof of our claim.

It follows from this key observation that in searching for concave Bernoulli functions that EU-rationalize $\mathcal{O}(\mathbf{e})$, we can confine our search to linear extensions of some $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$. But how do we check that \bar{u}_ℓ EU-rationalizes the data?

Let $\mathcal{N} = \{(a, b) \in \mathbb{R}_+^2 : a, b \in [0, \bar{r}] \text{ and either } a \text{ or } b \text{ is in } \mathcal{X}\}$. The set \mathcal{N} looks a like a net containing $\mathcal{G} = \mathcal{X}^2$. (In Figure 5, it is the net formed by the dotted lines.) If \bar{u}_ℓ EU-rationalizes $\mathcal{O}(\mathbf{e})$, then, obviously, at every observation t ,

$$(16) \quad \pi_1 \bar{u}_\ell(x_1^t) + \pi_2 \bar{u}_\ell(x_2^t) \geq \pi_1 \bar{u}_\ell(x_1) + \pi_2 \bar{u}_\ell(x_2)$$

$$\text{for all } \mathbf{x} = (x_1, x_2) \in \mathcal{N} \cap \partial \mathcal{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t).$$

(Recall that $\partial \mathcal{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t)$ is the upper boundary of the budget set $\mathcal{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t)$ and is equal to the budget line, i.e., $\{\mathbf{x} \in \mathbb{R}_+^2 : \mathbf{p}^t \cdot \mathbf{x} = e^t \mathbf{p}^t \cdot \mathbf{x}^t\}$.) It turns out that this condition is also *sufficient* for EU-rationalizability.

PROPOSITION 1: *Suppose the dataset $\mathcal{O}(\mathbf{e})$ is EU-rationalizable with probability $(\pi_1, \pi_2) \gg (0, 0)$ by a concave Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Let \mathcal{X} be a*

³⁶The definition of \mathcal{X}^{**} is similar to \mathcal{X}^* (see (5)), but the latter does not include \bar{r} .

³⁷Two points r and r' are adjacent in \mathcal{X} if there is no point in \mathcal{X} between r and r' .

³⁸Strictly speaking, \bar{h}_ℓ is not uniquely defined for $r > \bar{r}$, but the value of \bar{h}_ℓ beyond \bar{r} is irrelevant.

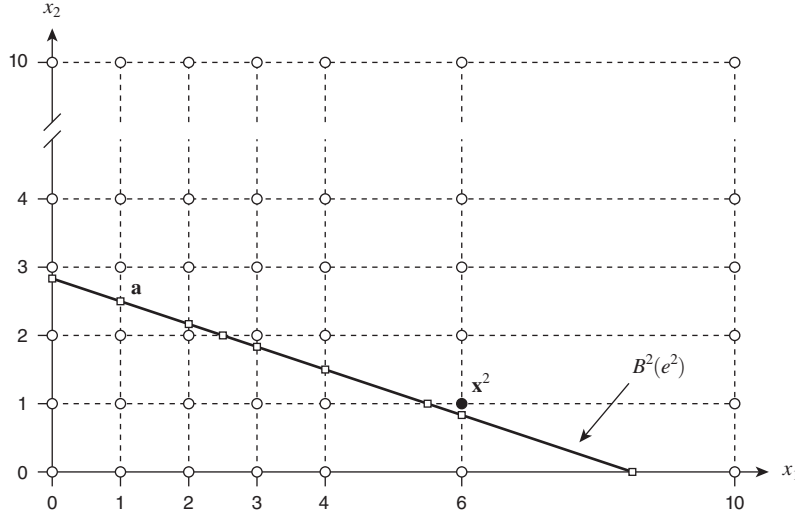


FIGURE 5. TEST FOR CONCAVE EU-RATIONALIZABILITY

finite set in \mathbb{R}_+ containing \mathcal{X}^{**} (as defined by (15)). Then the restriction of u to \mathcal{X} , $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$, has the following properties: (i) $\bar{u}(r) < \bar{u}(r')$ if $r < r'$; (ii) for any three adjacent points $r < r' < r''$ in \mathcal{X} ,

$$\frac{\bar{u}(r') - \bar{u}(r)}{r' - r} \geq \frac{\bar{u}(r'') - \bar{u}(r')}{r'' - r'};$$

and (iii) $\bar{u}_\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, the linear extension of \bar{u} , satisfies (16) at all t .

Conversely, if $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ satisfies (i), (ii), and (iii), then its linear extension \bar{u}_ℓ is a concave Bernoulli function that EU-rationalizes $\mathcal{O}(\mathbf{e})$.

If $\mathcal{O}(\mathbf{e})$ is EU-rationalizable by a concave Bernoulli function u , then (i) holds because u is increasing, (ii) holds because u is concave, and (iii) is necessary because the linear extension \bar{u}_ℓ also EU-rationalizes the data. For the converse, it is clear that conditions (i) and (ii) respectively guarantee that $\bar{u}_\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is an increasing and concave function. The proof in the Appendix shows that if (iii) also holds, then \bar{u}_ℓ EU-rationalizes $\mathcal{O}(\mathbf{e})$. The importance of this proposition is that it provides us with an easy-to-implement test, since conditions (i) to (iii) translate into a finite set of linear inequalities on a finite set of unknowns $\{\bar{u}(r)\}_{r \in \mathcal{X}}$, and checking whether a solution exists is a straightforward matter.

As an illustration of how this test works, we consider the dataset $\mathcal{O}(\mathbf{e})$ previously depicted in panel A of Figure 4. Given that the three observed choices are (2, 4), (6, 1), and (4, 3), and choosing $\bar{r} = 10$, we obtain $\mathcal{X}^{**} = \{0, 1, 2, 3, 4, 6, 10\}$. Letting $\mathcal{X} = \mathcal{X}^{**}$, the test involves setting up a collection of linear inequalities in the unknowns $\{\bar{u}(r)\}_{r \in \mathcal{X}}$ (corresponding to conditions (i) to (iii)) and checking if it has a solution. Conditions (i) and (ii) are clear enough, so let us explain condition (iii), which guarantees the optimality of the observed choice \mathbf{x}^t over a finite set of alternatives in $B^t(e^t)$. To be specific, consider its restrictions on the second observation. In Figure 5, we zoom in on $B^2(e^2)$, where \mathcal{N} is indicated by the dotted lines

dividing \mathbb{R}_+^2 . There are nine bundles in $\mathcal{N} \cap \partial\mathcal{B}^2(\mathbf{p}^2, e^2 \mathbf{p}^2 \cdot \mathbf{x}^2)$, indicated by the small squares on the budget line.³⁹ Condition (iii) requires that the expected utility of \mathbf{x}^2 , computed with \bar{u}_ℓ , be higher than the expected utility of those nine bundles. This translates into nine linear inequalities in the unknowns $\{\bar{u}(r)\}_{r \in \mathcal{X}}$. For example, the bundle $\mathbf{a} = (1, \lambda 2 + (1 - \lambda)3)$ for some $\lambda \in (0, 1)$ (which can easily be computed). The expected utility of \mathbf{a} is

$$\pi_1 \bar{u}_\ell(1) + \pi_2 \bar{u}_\ell(\lambda 2 + (1 - \lambda)3) = \pi_1 \bar{u}(1) + \pi_2 [\lambda \bar{u}(2) + (1 - \lambda) \bar{u}(3)],$$

since \bar{u}_ℓ is piecewise linear. Condition (iii) requires

$$\pi_1 \bar{u}(6) + \pi_2 \bar{u}(1) \geq \pi_1 \bar{u}(1) + \pi_2 \lambda \bar{u}(2) + \pi_2 (1 - \lambda) \bar{u}(3).$$

One could construct the remaining eight inequalities in a similar fashion.

IV. Implementation

We study the data collected from the well-known portfolio choice experiment in Choi et al. (2007), and from two other similar (more recent) experiments in Choi et al. (2014) and Halevy, Persitz, and Zrill (2018). The new tests developed in this paper allow us to evaluate, using a completely nonparametric approach, the empirical performance of different models of decision making under risk.

A. Model Performance

The experiment in Choi et al. (2007) was performed on 93 undergraduate subjects at the University of California, Berkeley. Every subject was asked to make consumption choices on 50 decision problems under risk. Each subject divided her budget between two Arrow-Debreu securities, with each security paying 1 token if the corresponding state was realized, and 0 otherwise. In a *symmetric treatment* applied to 47 subjects, each state of the world occurred with probability 1/2, and in a (balanced) *asymmetric treatment* applied to 46 subjects, the probabilities of the states were 1/3 and 2/3. These probabilities were objectively known. Lastly, income was normalized to 1, and the state prices were chosen at random and varied across rounds and subjects. In their analysis, Choi et al. (2007) first tested whether each subject's behavior is consistent with maximizing a locally nonsatiated utility function by performing a GARP test (or, strictly speaking, a modified version of the GARP test which characterizes rationalizability at a given (Afriat) cost efficiency threshold). Those subjects who passed GARP at a sufficiently high efficiency threshold were then fitted individually to a two-parameter version of the disappointment aversion model of Gul (1991).⁴⁰

The GRID method developed in this paper makes it possible to analyze the same data using purely revealed preference techniques. By applying the tests developed

³⁹Their coordinates can be easily computed from \mathcal{N} and $\partial\mathcal{B}^2(\mathbf{p}^2, e^2 \mathbf{p}^2 \cdot \mathbf{x}^2)$.

⁴⁰One parameter governed the distortion of state probabilities, and the other the degree of absolute/relative risk aversion.

in Sections I and II, we can calculate the Afriat or critical cost efficiency indices at which a subject's choice behavior is rationalizable by the expected utility (EU), disappointment aversion (DA), and rank-dependent utility (RDU) models. We can also do the same with the additional requirement that the Bernoulli function is concave, using the results in Section III and in online Appendix Section A4; we shall refer to these models as cEU, cDA, and cRDU.

It is well known that all of these models are contained within the larger class of *stochastically monotone* utility functions; these utility functions give strictly higher utility to the bundle \mathbf{x} compared to \mathbf{y} whenever \mathbf{x} first-order stochastically dominates \mathbf{y} (with respect to the objective state probabilities) and gives them the same utility whenever they are stochastically equivalent. In the Choi et al. (2007) experiment, there are just two states. In this case it is straightforward to check that when $\pi_1 = \pi_2 = 1/2$, a utility function is stochastically monotone if and only if it is strictly increasing and symmetric, and when $\pi_2 > \pi_1$, a utility function U is stochastically monotone if and only if it is strictly increasing and $U(a, b) > U(b, a)$ whenever $b > a$. Lastly, stochastically monotone utility functions are contained within the still larger class of locally nonsatiated utility functions.

Afriat's Theorem tells us that the rationalizability by locally nonsatiated utility maximization is observationally characterized by GARP. A test of stochastically monotone utility maximization was recently developed by Nishimura, Ok, and Quah (2017); this test has features similar to GARP and we shall refer to it as F-GARP (where "F" stands for first-order stochastic dominance). In both cases, it is also known that the axioms can be extended to test for rationalizability on modified datasets and can therefore be used to calculate the critical cost efficiency index (see Section II) at which the dataset is rationalizable. A detailed explanation of these axioms can be found in online Appendix Section A5.

To recap, for each subject in Choi et al. (2007), we calculate the critical cost efficiency index at which that subject is consistent with a given model. Altogether, there are eight models under consideration (locally nonsatiated utility maximization (GARP), stochastically monotone utility maximization (F-GARP), and RDU, DA, EU, cRDU, cDA, and cEU maximization). Therefore, to each subject under the asymmetric treatment, we assign eight efficiency indices (one for each model), while to each subject under the symmetric treatment, we assign six indices (since in the symmetric case, the RDU and DA models are identical, and the cRDU and cDA models are identical). When one model is, by definition, more stringent than another, its efficiency index must be weakly lower. So for a given subject, the efficiency index corresponding to GARP will be the highest, and the index corresponding to cEU will be the lowest. More generally, for each subject, the efficiency indices must be ordered the same way the models are nested, i.e.,

$$e_{cEU} \leq e_{EU}; \quad e_{cDA} \leq e_{DA}; \quad e_{cEU} \leq e_{cDA} \leq e_{cRDU} \leq e_{RDU};$$

$$e_{EU} \leq e_{DA} \leq e_{RDU} \leq e_{F-GARP} \leq e_{GARP}.$$

Basic Rationalizability.—Table 4 gives pass rates for the different models at three different thresholds of the Afriat or critical cost efficiency index: 0.9, 0.95, and 1,

TABLE 4—PASS RATES BY EFFICIENCY THRESHOLD (CHOI ET AL. 2007)

| | $\pi_1 = 1/2$ | | | | $\pi_1 \neq 1/2$ | | |
|--------|---------------|-------------|-------------|--------|------------------|-------------|------------|
| | $e = 0.90$ | $e = 0.95$ | $e = 1.00$ | | $e = 0.90$ | $e = 0.95$ | $e = 1.00$ |
| GARP | 38/47 (81%) | 32/47 (68%) | 12/47 (26%) | GARP | 37/46 (80%) | 29/46 (63%) | 4/46 (9%) |
| F-GARP | 30/47 (64%) | 23/47 (49%) | 1/47 (2%) | F-GARP | 33/46 (72%) | 26/46 (57%) | 3/46 (7%) |
| RDU | 30/47 (64%) | 23/47 (49%) | 1/47 (2%) | RDU | 33/46 (72%) | 24/46 (52%) | 2/46 (4%) |
| | | | | DA | 20/46 (43%) | 12/46 (26%) | 1/46 (2%) |
| EU | 30/47 (64%) | 18/47 (38%) | 1/47 (2%) | EU | 18/46 (39%) | 12/46 (26%) | 1/46 (2%) |
| cRDU | 24/47 (51%) | 12/47 (26%) | 0/47 (0%) | cRDU | 25/46 (54%) | 14/46 (30%) | 1/46 (2%) |
| | | | | cDA | 13/46 (28%) | 6/46 (13%) | 1/46 (2%) |
| cEU | 23/47 (49%) | 10/47 (21%) | 0/47 (0%) | cEU | 11/46 (24%) | 5/46 (11%) | 0/46 (0%) |

with the last corresponding to exact rationalizability.⁴¹ Across both treatments, 16 out of 93 subjects obey GARP exactly and are therefore consistent with locally non-satiated utility maximization, with subjects in the symmetric treatment performing distinctly better than those in the asymmetric treatment. Of the 16 subjects who pass GARP, only 4 pass F-GARP, and still fewer subjects are rationalizable by the more stringent models. Given that we observe 50 decisions for every subject, it is not altogether surprising that so many subjects should have violated GARP (let alone the more stringent conditions). The picture changes appreciably once we allow for some error in the form of cost inefficiencies: about 81 percent of the subjects pass GARP at efficiency thresholds exceeding 0.9, and 66 percent at thresholds exceeding 0.95, suggesting that a large fraction of the sample *does* indeed behave in a way that is broadly compatible with utility maximization.^{42,43}

In Figure 6, we depict (separately) the distributions of efficiency indices across subjects for 5 models under the symmetric treatment, and 6 models under the asymmetric treatment.⁴⁴ The models are nested by definition, so one would expect the efficiency distributions to be stacked, as indeed they are. In both panels, the topmost curve represents the distribution of efficiency indices corresponding to GARP (in other words, rationalizability by locally nonsatiated utility maximization), and the bottommost curve represents the distribution of indices corresponding to cEU maximization, which is the most stringent model.

The EU Model.—We can see from Table 4 that around one-half of all subjects passing GARP are then consistent with the EU model (at the 0.9 or 0.95

⁴¹ The efficiency indices corresponding to GARP were also calculated by Choi et al. (2007). The indices corresponding to all other models are new.

⁴² Furthermore, we know that the experiment provides a high-powered test of utility maximization, in the sense that we can safely dismiss the possibility that this outcome would have occurred randomly. Indeed, as Choi et al. (2007) has already pointed out, these pass rates are very different from what arises if one instead calculates efficiency indices for (uniformly) randomly generated budgetary data, following Bronars (1987); in that case, the proportion of synthetic (random) subjects passing GARP at efficiency thresholds exceeding 0.9 is very close to 0. (See Figure 4 in Choi et al. 2007.)

⁴³ In principle, a dataset can fail GARP exactly and yet have an efficiency index that equals 1, since the index is defined by a supremum (see Halevy, Persitz, and Zrill 2018 for an example of this phenomenon). However, these situations are nongeneric since they occur only when there are observations t and t' such that $\mathbf{p}' \cdot \mathbf{x}^t = \mathbf{p}^t \cdot \mathbf{x}^{t'}$. In Table 4, the subjects listed as having $e = 1$ also pass GARP exactly.

⁴⁴ We exclude the cRDU and cDA distributions in order to avoid congestion; the interested reader can find these distributions in online Appendix Section A6.3.

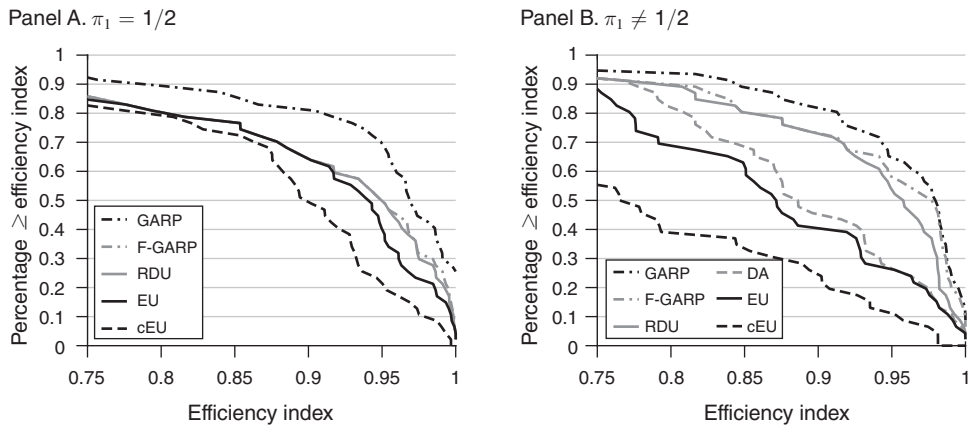


FIGURE 6. DISTRIBUTIONS OF EFFICIENCY INDICES (CHOI ET AL. 2007)

efficiency thresholds).⁴⁵ One might worry that the experimental design is insufficiently discriminating or powerful, so that, at a given efficiency threshold, random GARP-consistent datasets would have passed the EU test at the same rate, but this is far from the case. We can confirm that the pass rate for the EU model on a large collection of randomly generated GARP-consistent datasets is effectively zero. In fact, we can say even more. In Table 5, we report the results from a large collection of randomly generated datasets, all of which pass F-GARP at the given efficiency threshold (either 0.9 or 0.95).⁴⁶ At the 0.9 threshold, the pass rate for the EU model in these randomly generated data is 13 percent under the symmetric treatment and 0 percent under the asymmetric treatment; at the 0.95 threshold, the pass rates are effectively zero under both treatments. In other words, the observed EU pass rates are substantially higher than what would have arisen had the subjects been merely maximizing some stochastically monotone utility function. The cEU pass rates (where concavity is imposed on the Bernoulli function) are lower than the EU pass rates, but these too are substantially higher than the EU pass rates on randomly generated F-GARP-consistent data.⁴⁷

That said, it is worth emphasizing that there *is* a significant difference between the GARP and EU pass rates. Had the distributions for GARP and EU shown in Figure 6 been very close, we could have concluded that while subjects make mistakes when choosing from budget sets (since they fall short of consistency with basic rationality), they are nonetheless consistent with the EU model once that has been taken into account. However, since the distributions are distinct, that is not the case for a significant number of subjects.

We can have a sense of the *preference misspecification*, i.e., the extent to which the EU model misspecifies a subject's preference, by looking at the difference in

⁴⁵The precise ratios are $(30 + 18)/(38 + 37) = 64\%$ at the 0.9 threshold, and $(18 + 12)/(32 + 29) = 49\%$ at the 0.95 threshold.

⁴⁶Online Appendix Section A6.1 describes the procedure that we use to randomly generate GARP-consistent (or F-GARP-consistent) datasets at a given efficiency threshold.

⁴⁷Naturally, the cEU pass rates on randomly generated F-GARP-consistent data must be even lower than the EU pass rates reported in Table 5, since cEU is a more stringent model.

TABLE 5—PASS RATES FOR RANDOM F-GARP-CONSISTENT DATA (CHOI ET AL. 2007)

| | $\pi_1 = 1/2$ | | | $\pi_1 \neq 1/2$ | |
|-----|---------------|------------|-----|------------------|------------|
| | $e = 0.90$ | $e = 0.95$ | | $e = 0.90$ | $e = 0.95$ |
| RDU | 25% | 1% | RDU | 1% | 0% |
| | | | DA | 0% | 0% |
| EU | 13% | 1% | EU | 0% | 0% |

the efficiency index between GARP and the EU model.⁴⁸ The median difference is 0.027 for subjects in the symmetric treatment and 0.075 for those in the asymmetric treatment. It exceeds 0.05 for 17 out of 47 subjects in the symmetric treatment and 30 out of 46 subjects in the asymmetric treatment.

The RDU and DA Models.—Can these models play a useful role in explaining behavior which is not captured by the EU model? Under the symmetric treatment, there is little scope for these models to capture the behavior of subjects not already compatible with the EU model, since the latter is already accommodating most of the subjects who pass F-GARP. (Note the closeness of the F-GARP and EU distributions in panel A of Figure 6.)

However, the RDU model appears to capture the behavior of subjects more successfully than the EU model under the asymmetric treatment, covering almost all F-GARP-consistent behavior (see Table 4 or panel B of Figure 6).⁴⁹ When we require the Bernoulli function to be concave, the rank-dependent utility model can no longer account for nearly all F-GARP-consistent behavior, but even then it captures many more subjects than expected utility (compare cRDU with cEU). On the other hand, this is not true of the DA model. Under both the symmetric and asymmetric treatments, the DA pass rates are only slightly higher than the EU pass rates, and the same is true when comparing cDA with cEU.

Table 6 summarizes these observations. We record (as a fraction of all subjects within each treatment) the pass rates for the cEU and EU models. We also report the marginal contributions of the RDU and DA models (relative to EU or cEU) in explaining the data. For example, under the asymmetric treatment, at the 0.9 threshold, 15 subjects out of 46 pass RDU but fail EU (see the row beginning RDU\EU); using this information, we can form a 95 percent binomial proportion confidence interval on the probability that a subject is rationalizable by RDU but not by EU, which turns out to be [0.195, 0.480].⁵⁰

We conduct similar tests on the data collected by Choi et al. (2014) and Halevy, Persitz, and Zrill (2018). In both experiments, subjects allocated investment between

⁴⁸We are broadly following Halevy, Persitz, and Zrill (2018) in using this measure of preference misspecification. Halevy, Persitz, and Zrill (2018) compares the change in the money metric index (which is essentially Varian's inconsistency index) between GARP and a parametric version of the DA model and interpret that difference as a measure of the misspecification.

⁴⁹We discuss the probability distortions which are required in order to rationalize these datasets in online Appendix Section A6.2.

⁵⁰All confidence intervals in this table are exact, and calculated using the Clopper-Pearson method.

TABLE 6—CONFIDENCE INTERVALS ON PREFERENCE TYPES (CHOI ET AL. 2007)

| | $\pi_1 = 1/2$ | | | |
|------------|---------------|----------------|---------------|----------------|
| | $e = 0.90$ | | $e = 0.95$ | |
| | Sample prop. | Conf. interval | Sample prop. | Conf. interval |
| cEU | 23/47 (0.489) | [0.341, 0.639] | 10/47 (0.213) | [0.107, 0.357] |
| EU | 30/47 (0.638) | [0.485, 0.773] | 18/47 (0.383) | [0.245, 0.536] |
| EU \ cEU | 7/47 (0.149) | [0.062, 0.283] | 8/47 (0.170) | [0.076, 0.308] |
| RDU \ EU | 0/47 (0.000) | [0.000, 0.075] | 5/47 (0.106) | [0.035, 0.231] |
| cRDU \ cEU | 1/47 (0.021) | [0.001, 0.113] | 2/47 (0.043) | [0.005, 0.145] |

| | $\pi_1 \neq 1/2$ | | | |
|------------|------------------|----------------|---------------|----------------|
| | $e = 0.90$ | | $e = 0.95$ | |
| | Sample prop. | Conf. interval | Sample prop. | Conf. interval |
| cEU | 11/46 (0.239) | [0.126, 0.388] | 5/46 (0.109) | [0.036, 0.236] |
| EU | 18/46 (0.391) | [0.251, 0.546] | 12/46 (0.261) | [0.143, 0.411] |
| EU \ cEU | 7/46 (0.152) | [0.063, 0.289] | 7/46 (0.152) | [0.063, 0.289] |
| DA \ EU | 2/46 (0.043) | [0.005, 0.148] | 0/46 (0.000) | [0.000, 0.077] |
| RDU \ EU | 15/46 (0.326) | [0.195, 0.480] | 12/46 (0.261) | [0.143, 0.411] |
| cDA \ cEU | 2/46 (0.043) | [0.005, 0.148] | 1/46 (0.022) | [0.001, 0.115] |
| cRDU \ cEU | 14/46 (0.304) | [0.177, 0.458] | 9/46 (0.196) | [0.094, 0.339] |

two Arrow-Debreu securities, with commonly known equiprobable states. Thus, the designs closely resembled the symmetric treatment in Choi et al. (2007).⁵¹

Analysis of the Choi et al. (2014) Dataset.—This experiment was conducted on 1,182 CentERpanel adult members, where the latter is meant to be representative of the Dutch-speaking population of the Netherlands. Each subject made allocation decisions on 25 linear budget sets; since this is just one-half of the number of decisions in Choi et al. (2007), the pass rates should be higher if the subject population is the same, but it is not. As has already been noted by Choi et al. (2014), the pass rates for GARP at any efficiency threshold are instead *lower* those in Choi et al. (2007). This observation can now be extended further: the pass rates are across the board lower for all tests, and not just for GARP. This is clear if we compare Figure 7 with panel A of Figure 6. That said, it is also clear from these two figures that certain qualitative features of the data are the same in both experiments. In particular, around one-half of the subjects who pass GARP at a given efficiency threshold are also consistent with the EU model. There is a significant difference in the pass rates between GARP and F-GARP, and the EU model manages to explain a very large share of subjects who pass F-GARP. In this experiment (but unlike in Choi et al. 2007), the cEU model also manages to account for many subjects who pass F-GARP. Since the rank-dependent utility (equivalently, disappointment aversion) model is more stringent than F-GARP, the model's contribution, as measured by the proportion of subjects who obey RDU but not EU, or cRDU but not cEU (at some reasonable threshold), is modest. This echoes our finding for the symmetric treatment in Choi et al. (2007).

⁵¹ Since the RDU and DA models coincide when states are equiprobable, the interesting distinction that we find between them in the asymmetric treatment in Choi et al. (2007) could not be further investigated. We think there is a case for including asymmetric treatments in future experiments, or even to have the same subject choosing under different state probabilities.

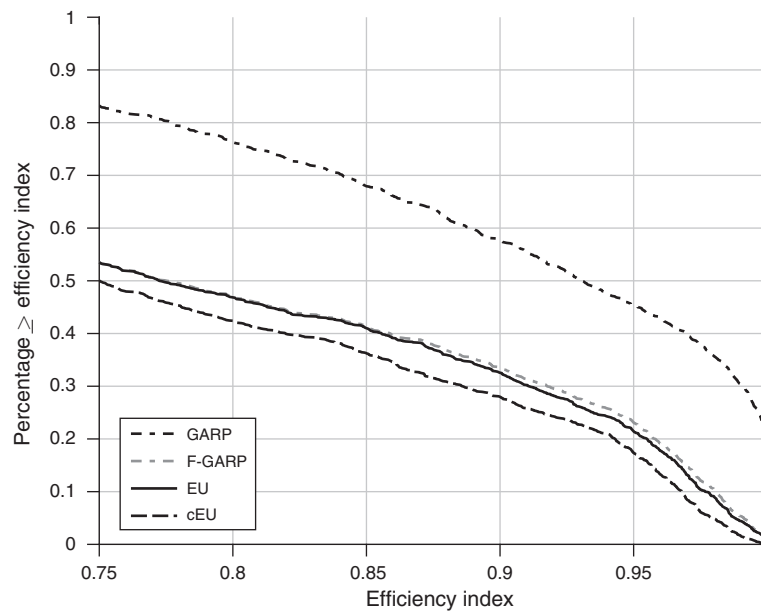


FIGURE 7. DISTRIBUTIONS OF EFFICIENCY INDICES (CHOI ET AL. 2014)

More details of our analysis of the Choi et al. (2014) data can be found in online Appendix Section A7. We also explore in that section the relationship between a subject's efficiency indices (for different models) and various socioeconomic variables and outcomes (such as age, education, and wealth), extending the analysis in Choi et al. (2014).

Analysis of the Halevy, Persitz, and Zrill (2018) Dataset.—This experiment was conducted on 207 primarily undergraduate subjects at the University of British Columbia, with a set of portfolio choice problems forming the first part of a two-part experiment. (We discuss the second part of the experiment in Section IVB.) Each subject made allocation decisions on 22 linear budget sets; since this number is lower than in Choi et al. (2007) (where each subject made decisions on 50 budget sets) and the sample population is similar, one would expect the pass rates in Halevy, Persitz, and Zrill (2018) to be generally higher than those in Choi et al. (2007), and this is what we find. However, the relative performance of the different models (relative to one another) is broadly similar across the two experiments. The EU model performs well: significantly more than half of the subjects who pass GARP (at some reasonable efficiency threshold) also pass the test for the EU model. That said, there is still a distinct difference in performance between the models; indeed a significant number of subjects who pass GARP fail F-GARP (and thus EU). The RDU model makes only a modest contribution relative to EU, and similarly, consistency with the cRDU model is only slightly higher than with the cEU model. These observations are clear in Figure 8, which depicts the distributions of efficiency indices for different models in this dataset.

While parametric models are easy to use and have other advantages, they will by definition fit a dataset less well than their nonparametric counterparts, so there is

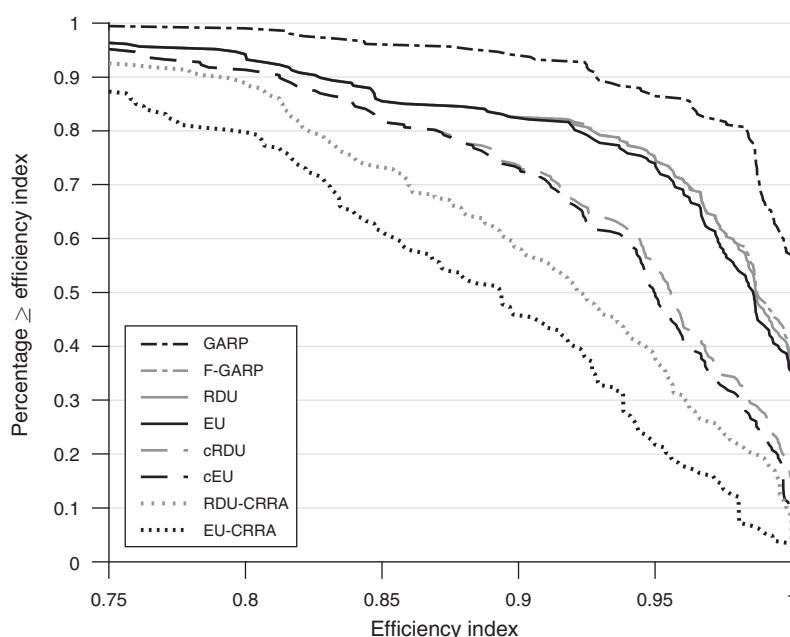


FIGURE 8. DISTRIBUTIONS OF EFFICIENCY INDICES (HALEVY, PERSITZ, AND ZRILL 2018)

some advantage in developing procedures to assess the size of that loss of fit. Halevy, Persitz, and Zrill (2018) evaluated the basic rationality of their subjects (through GARP) and also the goodness-of-fit of parametric versions of the rank-dependent and expected utility models. Following their example, we calculate the efficiency indices for the rank-dependent and expected utility models, with the Bernoulli functions confined to the CRRA class.⁵² Their distributions are depicted in Figure 8 (see the RDU-CRRA and EU-CRRA curves).

Two things are clear from this exercise. First, the misspecification involved in using a parametric form appears to be substantial. This is suggested by Figure 8 where the pass rates for GARP are much higher than that for RDU-CRRA or EU-CRRA.⁵³ We can also measure the size of the preference misspecification by the difference in the efficiency index between GARP and RDU-CRRA for each subject. This difference exceeds 0.05 for 118 (out of 207) subjects and exceeds 0.10 for 76 subjects. These observations are broadly in line with those made by Halevy,

⁵²The algorithm for calculating the critical cost efficiency index for the EU-CRRA and RDU-CRRA models is straightforward and does not involve the GRID method. Consider, for example, the EU-CRRA case. At the efficiency vector $\mathbf{e} = (e, e, \dots, e)$ for $e \in (0, 1]$, we can determine if the modified dataset $\mathcal{O}(\mathbf{e})$ is consistent with EU-CRRA for a given coefficient of relative risk aversion η . We denote the EU-CRRA utility function of the bundle $\mathbf{x} \in \mathbb{R}_+^2$ by $U(\mathbf{x}; \eta)$ and the EU-CRRA indirect utility at price \mathbf{p} and income m by $V(\mathbf{p}, m; \eta)$ (the formula for which can be easily calculated). Note that $\mathcal{O}(\mathbf{e})$ is rationalized by $U(\cdot; \eta)$ if and only if $U(\mathbf{x}^t; \eta) \geq V(\mathbf{p}^t, e\mathbf{p}^t \cdot \mathbf{x}^t; \eta)$ for $t = 1, 2, \dots, T$. This can be checked for a given η , and by letting η take different values we can establish if $\mathcal{O}(\mathbf{e})$ is EU-CRRA-rationalizable for a given \mathbf{e} . Lastly, we perform a binary search over $(0, 1]$ in order to determine the critical value of e , as described in footnote 34.

⁵³For example, at any efficiency threshold of 0.9, more than 90 percent of subjects pass GARP but fewer than 60 percent are consistent with RDU-CRRA.

Persitz, and Zrill (2018).⁵⁴ Second, this misspecification is considerably worse than that for the corresponding *nonparametric* model; indeed, the difference in the efficiency index between GARP and cRDU exceeds 0.05 for 82 (out of 207) subjects and exceeds 0.10 for 38 subjects.

Online Appendix Section A8 provides further analysis of the data collected by Halevy, Persitz, and Zrill (2018). Note that Halevy, Persitz, and Zrill (2018) makes use of Varian's efficiency index, the calculation of which is feasible in the case of GARP and the parametric models they consider. We have not used Varian's index because its calculation for *all* of the nonparametric models that we consider is too computationally demanding. We did, however, calculate Varian's index (exactly) for GARP and F-GARP and also have good approximations for the EU model. In all three cases, the Afriat and Varian indices are highly correlated. Our Varian indices are obtained via a new search algorithm (see online Appendix Section A9) that may be of independent interest.

Comparison with Empirical Findings in Other Papers.—There is a large empirical literature that evaluates the performance of different models of choice under risk using experimental or field data, and our results appear to be broadly in line with the findings obtained in earlier studies, even though the very different empirical methods employed make formal comparisons difficult. In particular, other papers have concluded that the rank-dependent utility model performs well (see, for example, Bruhin, Fehr-Duda, and Epper 2010 and Barseghyan et al. 2013 and their references), which is something that we also observe, at least in the asymmetric treatment in Choi et al. (2007). We find that the expected utility model captures a significant number of subjects, though by no means everyone, which is broadly consistent with the not altogether uncommon finding that this canonical model puts in a respectable performance (see, for example, Hey and Orme 1994). Lastly, the relatively poor performance of the disappointment aversion model has also been noted in some other studies such as Hey and Orme (1994) and Barseghyan et al. (2013).

B. Out-of-Sample Predictions

We use the data from Choi et al. (2007) and Halevy, Persitz, and Zrill (2018) to make out-of-sample predictions. For different reasons, neither experiment is ideal for our purpose, but these applications are still indicative about the potential usefulness of our procedures.

The Halevy, Persitz, and Zrill (2018) Data.—In the second part of the experiment in Halevy, Persitz, and Zrill (2018), each subject was asked to make a number of choices between two lotteries: lottery A, which pays off some amount for sure, and lottery B, which has two unequal but equiprobable payoffs.⁵⁵ The objective was to adjudicate between two competing parametric recovery approaches, one using

⁵⁴ Using Varian's index rather than Afriat's, Halevy, Persitz, and Zrill (2018) reaches a similar conclusion that the contribution to inconsistency from parametric misspecification is large relative to that from the failure of basic rationality.

⁵⁵ For example, lottery A might pay 50 for sure and lottery B might pay 60 and 40 with equal probability.

nonlinear least squares (NLLS) and another using a money metric index (MMI) (which is essentially Varian's efficiency index). The recovery exercise first involved fitting (using either procedure) the RDU-CRRA model to each subject's portfolio choice data (from the first part of the experiment). This yielded two parameters for each subject: β , which governs the probability distortion (see (13)), and η , the coefficient of relative risk aversion. With these parameters, one could then predict the subject's choice between any two lotteries A and B. The lotteries A and B were not randomly chosen, but instead tailor-made to each subject so that the two recovery methods would lead to *different predictions* for each pair.

Our nonparametric approach could also be used to predict lottery choices. First, for a given model, say RDU, we calculate a subject's efficiency index. Suppose the index is e^* ; then we test whether a preference for A over B is consistent with the RDU model given the dataset $\{(\mathbf{x}^t, B^t(e^*))\}_{t=1}^T$ (in the sense explained in Section IC). Since $\{(\mathbf{x}^t, B^t(e^*))\}_{t=1}^T$ is RDU-rationalizable by definition, either a preference for A over B or its reverse will be consistent. It is also possible that both are consistent, in which case the model cannot discriminate between these lotteries; note that this is where a nonparametric model differs from a parametric model, since the recovered parameters are typically unique in the latter (which lead to unique predictions).

In the experiment, each subject (of 207) was asked to make 9 pairwise choices, giving 1,863 binary comparisons in total. Using the RDU model, we find that only 79 of these cases (around 4 percent) are discriminating; among these 79 cases, 48 are correctly predicted, about 61 percent. Under cRDU, discrimination is sharpened, with 957 (51 percent) being discriminating, and among these, 508 (53 percent) are correctly predicted. Lastly, under RDU-CRRA, which is a parametric model, all 1,863 binary comparisons are discriminating and 1,051 of these (56 percent) are correctly predicted;⁵⁶ this result is in line with Halevy, Persitz, and Zrill (2018), which correctly predicts 54 percent of 1,827 binary choices when the RDU-CRRA parameters are recovered using the MMI method (which means, given the special way lotteries were chosen in that experiment, that the NLLS method is correct for 46 percent of the choices).^{57,58}

The Choi et al. (2007) Data.—The special way in which lotteries are chosen in the Halevy, Persitz, and Zrill (2018) experiment gives us only limited information on whether our nonparametric procedure is discriminating enough to be useful for making out-of-sample predictions. To explore this issue further, we conduct a simple but instructive exercise using the data collected from the symmetric treatment in Choi et al. (2007). We first identify those subjects with RDU efficiency indices exceeding 0.9. For each subject, we choose an observation at random; call it t' . We then randomly choose a bundle \mathbf{y} that is undominated⁵⁹ by $\mathbf{x}^{t'}$ and satisfies $\mathbf{p}' \cdot \mathbf{y} = 0.9 \mathbf{p}' \cdot \mathbf{x}^{t'}$. We then ask whether we could “predict” the choice between $\mathbf{x}^{t'}$ and \mathbf{y} from the remaining 49 portfolio choice observations (using the

⁵⁶ In a sense, RDU-CRRA is both correct more often and wrong more often than cRDU.

⁵⁷ Notice that 4 subjects were removed in Halevy, Persitz, and Zrill (2018), giving a sample of 203 subjects.

⁵⁸ Since the lottery pairs in this experiment were chosen in a very special way, the reader should take care not to conclude that these methods, whether parametric or otherwise, make correct predictions only one-half of the time when presented with a “random” pair of lotteries.

⁵⁹ By undominated, we mean that if $\mathbf{y} = (a, b)$ then $\mathbf{x}^{t'} \not\geq (a, b)$ and $\mathbf{x}^{t'} \not\geq (b, a)$.

procedure set out in Section IC). Since we have chosen subjects with efficiency indices above 0.9, a preference for \mathbf{x}' over \mathbf{y} *must* be consistent with the 49 portfolio choice decisions. The issue is whether a preference for \mathbf{y} over \mathbf{x}' is *also* consistent with the 49 observations; if so, it means that the nonparametric procedure has failed to be discriminating “out of sample.”

Under RDU, 30 of 47 subjects have efficiency indices exceeding 0.9. For each of these subjects, we perform 2 independent (i.e., drawn from possibly different budgets) predictive exercises and find that 50 of 60 (83 percent) of these are predictively discriminating. In the case of cRDU, 24 of 47 subjects have efficiency indices exceeding 0.9 and we find that 46 of 48 (96 percent) of the predictive exercises are discriminating. Obviously this simple exercise is no more than indicative, but it does suggest that our nonparametric procedure is capable of making sharp predictions out of sample.

APPENDIX

PROOF OF THEOREM 1:

We require the following lemma.

LEMMA 1: *Let $\{C^t\}_{t=1}^T$ be a finite collection of constraint sets in $\mathbb{R}_+^{\bar{s}}$ that are compact and downward closed (i.e., if $\mathbf{x} \in C^t$ then so is $\mathbf{y} \in \mathbb{R}_+^{\bar{s}}$ such that $\mathbf{y} < \mathbf{x}$) and let the functions $\{\phi(\cdot, t)\}_{t=1}^T$ be continuous and increasing in all dimensions. Suppose that there is a finite set \mathcal{X} of \mathbb{R}_+ , a strictly increasing function $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$, and $\{M^t\}_{t=1}^T$ such that the following holds:*

$$(A1) \quad M^t \geq \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \quad \text{for all } \mathbf{x} \in C^t \cap \mathcal{G}$$

and

$$(A2) \quad M^t > \phi(\bar{\mathbf{u}}(\mathbf{x}), t) \quad \text{for all } \mathbf{x} \in (C^t \setminus \partial C^t) \cap \mathcal{G},$$

where $\mathcal{G} = \mathcal{X}^{\bar{s}}$ and $\bar{\mathbf{u}}(\mathbf{x}) = (\bar{u}(x_1), \bar{u}(x_2), \dots, \bar{u}(x_{\bar{s}}))$. Then there is a Bernoulli function $u : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that extends \bar{u} such that

$$(A3) \quad M^t \geq \phi(\mathbf{u}(\mathbf{x}), t) \quad \text{for all } \mathbf{x} \in C^t$$

and

$$(A4) \quad \text{if } \mathbf{x} \in C^t \text{ and } M^t = \phi(\bar{\mathbf{u}}(\mathbf{x}), t), \text{ then } \mathbf{x} \in \partial C^t \cap \mathcal{G} \text{ and } M^t = \phi(\mathbf{u}(\mathbf{x}), t).$$

Remark 1: The property (A4) needs some explanation. Conditions (A1) and (A2) allow for the possibility that $M^t = \phi(\bar{\mathbf{u}}(\mathbf{x}'), t)$ for some $\mathbf{x}' \in \partial C^t \cap \mathcal{G}$; we denote the set of points in $\partial C^t \cap \mathcal{G}$ with this property by X' . Clearly any extension u will preserve this property, i.e., $M^t = \phi(\mathbf{u}(\mathbf{x}'), t)$ for all $\mathbf{x}' \in X'$. Property (A4) says that we can choose u such that for all $\mathbf{x} \in C^t \setminus X'$, we have $M^t > \phi(\mathbf{u}(\mathbf{x}), t)$.

PROOF:

We shall prove this result by induction on the dimension of the space containing the constraint sets. It is trivial to check that the claim is true if $\bar{s} = 1$. In this case, \mathcal{G} consists of a finite set of points on \mathbb{R}_+ and each C^t is a closed interval with 0 as its minimum. Now let us suppose that the claim holds for $\bar{s} = m$ and we shall prove it for $\bar{s} = m + 1$. If, for each t , there is a strictly increasing and continuous utility function $u^t : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ extending \bar{u} such that (A3) and (A4) hold, then the same conditions will hold for the increasing and continuous function $u = \min_t u^t$. So we can focus our attention on constructing u^t for a single constraint set C^t .

Suppose $\mathcal{X} = \{0, r^1, r^2, r^3, \dots, r^I\}$, with $r^0 = 0 < r^i < r^{i+1}$, for $i = 1, 2, \dots, I-1$. Let $\bar{r} = \max\{r \in \mathbb{R}_+ : (r, 0, 0, \dots, 0) \in C^t\}$ and suppose that $(r^i, 0, 0, \dots, 0) \in C^t$ if and only if $i \leq N$ (for some $N \leq I$). Consider the collection of sets of the form $D^i = \{\mathbf{y} \in \mathbb{R}_+^m : (r^i, \mathbf{y}) \in C^t\}$ (for $i = 1, 2, \dots, N$); this is a finite collection of compact and downward closed sets in \mathbb{R}_+^m . By the induction hypothesis applied to $\{D^i\}_{i=1}^N$, with $\{\phi(\bar{u}(r^i), \cdot, t)\}_{i=1}^N$ as the collection of functions, there is a strictly increasing function $u^* : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ extending \bar{u} such that

$$(A5) \quad M^t \geq \phi(\bar{u}(r^i), \mathbf{u}^*(\mathbf{y}), t) \quad \text{for all } (r^i, \mathbf{y}) \in C^t$$

and

$$(A6) \quad \text{if } (r^i, \mathbf{y}) \in C^t \quad \text{and} \quad M^t = \phi(\bar{u}(r^i), \mathbf{u}^*(\mathbf{y}), t),$$

$$\text{then } (r^i, \mathbf{y}) \in \partial C^t \cap \mathcal{G} \quad \text{and} \quad M^t = \phi(\bar{u}(r^i, \mathbf{y}), t).$$

For each $r \in [0, \bar{r}]$, define

$$U(r) = \left\{ u \leq u^*(r) : \max\left\{ \phi(u, \mathbf{u}^*(\mathbf{y}), t) : (r, \mathbf{y}) \in C^t \right\} \leq M^t \right\}.$$

This set is nonempty; indeed $\bar{u}(r^k) = u^*(r^k) \in U(r)$, where r^k is the largest element in \mathcal{X} that is weakly smaller than r . This is because, if $(r, \mathbf{y}) \in C^t$ then so is (r^k, \mathbf{y}) , and (A5) guarantees that $\phi(\bar{u}(r^k), \mathbf{u}^*(\mathbf{y}), t) \leq M^t$. The downward closedness of C^t and the fact that u^* is increasing also guarantees that $U(r) \subseteq U(r')$ whenever $r < r'$. Now define $\tilde{u}(r) = \sup U(r)$; the function \tilde{u} has a number of significant properties. (i) For $r \in \mathcal{X}$, $\tilde{u}(r) = u^*(r) = \bar{u}(r)$ (by the induction hypothesis). (ii) \tilde{u} is a nondecreasing function since U is nondecreasing. (iii) $\tilde{u}(r) > \bar{u}(r^k)$ if $r > r^k$, where r^k is largest element in \mathcal{X} smaller than r . Indeed, because C^t is compact and ϕ continuous, $\phi(\tilde{u}(r), \mathbf{u}^*(\mathbf{y}), t) \leq M^t$ for all $(r, \mathbf{y}) \in C^t$. By way of contradiction, suppose $\tilde{u}(r) = \bar{u}(r^k)$ and hence $\tilde{u}(r) < u^*(r)$. It follows from the definition of $\tilde{u}(r)$ that, for any sequence u_n , with $\tilde{u}(r) < u_n < u^*(r)$ and $\lim_{n \rightarrow \infty} u_n = \tilde{u}(r)$, there is $(r, \mathbf{y}_n) \in C^t$ such that $\phi(u_n, \mathbf{u}^*(\mathbf{y}_n), t) > M^t$. Since C^t is compact, we may assume with no loss of generality that $\mathbf{y}_n \rightarrow \hat{\mathbf{y}}$ and $(r, \hat{\mathbf{y}}) \in C^t$, from which we obtain $\phi(\tilde{u}(r), \mathbf{u}^*(\hat{\mathbf{y}}), t) = M^t$. Since C^t is downward closed, $(r^k, \hat{\mathbf{y}}) \in C^t$ and, since $\bar{u}(r^k) = u^*(r^k)$, we have $\phi(u^*(r^k), \mathbf{u}^*(\hat{\mathbf{y}}), t) = M^t$. This can only occur if $(r^k, \hat{\mathbf{y}}) \in \partial C^t \cap \mathcal{G}$ (because of (A6)), but it is clear that $(r^k, \hat{\mathbf{y}}) \notin \partial C^t$ since $(r^k, \hat{\mathbf{y}}) < (r, \hat{\mathbf{y}})$. (iv) If $r_n < r^i$ for all n and $r_n \rightarrow r^i \in \mathcal{X}$, then $\tilde{u}(r_n) \rightarrow u^*(r^i)$. Suppose to the contrary, that the limit is $\hat{u} < u^*(r^i) = \bar{u}(r^i)$. Since u^* is continuous, we

can assume, without loss of generality, that $\tilde{u}(r_n) < u^*(r_n)$. By the compactness of C^t , the continuity of ϕ , and the definition of \tilde{u} , there is $(r_n, \mathbf{y}_n) \in C^t$ such that $\phi(\tilde{u}(r_n), \mathbf{u}^*(\mathbf{y}_n), t) = M^t$. This leads to $\phi(\tilde{u}, \mathbf{u}^*(\mathbf{y}'), t) = M^t$, where \mathbf{y}' is an accumulation point of \mathbf{y}_n and $(r^i, \mathbf{y}') \in C^t$. But since ϕ is strictly increasing, we obtain $\phi(u^*(r^i), \mathbf{u}^*(\mathbf{y}'), t) > M^t$, which contradicts (A5).

Given the properties of \tilde{u} , we can find a continuous and strictly increasing function u^t such that u^t extends \bar{u} , i.e., $u^t(r) = \bar{u}(r)$ for $r \in \mathcal{X}$, $u^t(r) < u^*(r)$ for all $r \in \mathbb{R}_+ \setminus \mathcal{X}$ and $u^t(r) < \tilde{u}(r) \leq u^*(r)$ for all $r \in [0, \bar{r}] \setminus \mathcal{X}$. (In fact we can choose u^t to be smooth everywhere except possibly on \mathcal{X} .) We claim that (A3) and (A4) are satisfied for C^t . To see this, note that for $r \in \mathcal{X}$ and $(r, \mathbf{y}) \in C^t$, the induction hypothesis guarantees that (A5) and (A6) hold and they will continue to hold if u^* is replaced by u^t . In the case where $r \notin \mathcal{X}$ and $(r, \mathbf{y}) \in C^t$, since $u^t(r) < \tilde{u}(r)$ and ϕ is increasing, we obtain $M^t > \phi(u^t(r, \mathbf{y}), t)$. ■

PROOF OF THEOREM 1:

This follows immediately from Lemma 1 if we set $C^t = \underline{B}^t$, and $M^t = \phi(\bar{\mathbf{u}}(\mathbf{x}^t), t)$. If \bar{u} obeys conditions (6) and (7) then it obeys conditions (A1) and (A2). The rationalizability of \mathcal{O} by $\{\phi(\cdot, t)\}_{t \in T}$ then follows from (A3). ■

PROOF OF PROPOSITION 1 (Sufficiency):

Suppose there is $\bar{u} : \mathcal{X} \rightarrow \mathbb{R}_+$ satisfying (i) to (iii). Then (i) guarantees that \bar{u}_ℓ is strictly increasing on \mathbb{R}_+ and (ii) guarantees that \bar{u}_ℓ is concave. We claim that with \bar{u}_ℓ as the Bernoulli function, \mathbf{x}^t has higher expected utility than any bundle in $B^t(e^t)$. By definition, \bar{u}_ℓ is linear between adjacent values of \mathcal{X} ; it follows that the map from (a, b) to its expected utility $\pi_1 \bar{u}_\ell(a) + \pi_2 \bar{u}_\ell(b)$ is also linear for all $(a, b) \in [r, r'] \times [m, m']$, where r and r' are adjacent points in \mathcal{X} (and similarly m and m'). A linear map is maximized at an extreme point; thus if $([r, r'] \times [m, m']) \cap \partial \mathcal{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t)$ is nonempty then there is a bundle (a^*, b^*) maximizing expected utility in this set with either $a^* \in \{r, r'\}$ or $b^* \in \{m, m'\}$. More generally, there must be a bundle (a^{**}, b^{**}) that maximizes expected utility in $\mathcal{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t)$ and is contained in $\partial \mathcal{B}(\mathbf{p}^t, e^t \mathbf{p}^t \cdot \mathbf{x}^t) \cap \mathcal{N}$. It follows that (16) is sufficient to guarantee the optimality of \mathbf{x}^t in $B^t(e^t)$. ■

REFERENCES

- Afriat, Sydney. 1967. "The Construction of Utility Functions from Expenditure Data." *International Economic Review* 8 (1): 67–77.
- Afriat, Sydney. 1972. "Efficiency Estimation of Production Function." *International Economic Review* 13 (3): 568–98.
- Afriat, Sydney. 1973. "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method." *International Economic Review* 14 (2): 460–72.
- Ahn, David, Syngjoo Choi, Douglas Gale, and Shachar Kariv. 2014. "Estimating Ambiguity Aversion in a Portfolio Choice Experiment." *Quantitative Economics* 5 (2): 195–223.
- Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (2): 737–53.
- Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *American Economic Review* 102 (7): 3333–56.
- Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2013. "The Nature of Risk Preferences: Evidence from Insurance Choices." *American Economic Review* 103 (6): 2499–2529.

- Bayer, Ralph, Subir Bose, Matthew Polisson, and Ludovic Renou. 2013. "Ambiguity Revealed." IFS Working Paper W13/05.
- Beatty, Timothy K. M., and Ian A. Crawford. 2011. "How Demanding Is the Revealed Preference Approach to Demand?" *American Economic Review* 101 (6): 2782–95.
- Bronars, Stephen G. 1987. "The Power of Nonparametric Tests of Preference Maximization." *Econometrica* 55 (3): 693–98.
- Brown, Donald J., and Rosa L. Matzkin. 1996. "Testable Restrictions on the Equilibrium Manifold." *Econometrica* 64 (6): 1249–62.
- Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper. 2010. "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion." *Econometrica* 78 (4): 1375–1412.
- Bussieck, Michael R., and Stefan Vigerske. 2010. "MINLP Solver Software." In *Wiley Encyclopedia of Operations Research and Management Science*, edited by James J. Cochran et al. New York: Wiley.
- Cappelen, Alexander W., Shachar Kariv, Erik Ø. Sørensen, and Bertil Tungodden. 2015. "Is There a Development Gap in Rationality?" Unpublished.
- Chambers, Christopher P., Federico Echenique, and Kota Saito. 2016. "Testing Theories of Financial Decision Making." *Proceedings of the National Academy of Sciences* 113 (15): 4003–08.
- Chambers, Christopher P., Ce Liu, and Seung-Keun Martinez. 2016. "A Test for Risk-Averse Expected Utility." *Journal of Economic Theory* 163: 775–85.
- Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv. 2007. "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review* 97 (5): 1921–38.
- Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who Is (More) Rational?" *American Economic Review* 104 (6): 1518–50.
- Diewert, Walter E. 1973. "Afriat and Revealed Preference Theory." *Review of Economic Studies* 40 (3): 419–25.
- Diewert, Walter E. 2012. "Afriat's Theorem and some Extensions to Choice under Uncertainty." *Economic Journal* 122 (560): 305–31.
- Dybvig, Philip, and Heraklis Polemarchakis. 1981. "Recovering Cardinal Utility." *Review of Economic Studies* 48 (1): 159–66.
- Dziwulski, Pawel, and John K.-H. Quah. 2014. "Testing for Production with Complementarities." University of Oxford Department of Economics Discussion Paper 722.
- Echenique, Federico, Sangmok Lee, and Matthew Shum. 2011. "The Money Pump as a Measure of Revealed Preference Violations." *Journal of Political Economy* 119 (6): 1201–23.
- Echenique, Federico, and Kota Saito. 2015. "Savage in the Market." *Econometrica* 83 (4): 1467–95.
- Epstein, Larry G. 2000. "Are Probabilities Used in Markets?" *Journal of Economic Theory* 91 (1): 86–90.
- Fishburn, Peter C. 1975. "Separation Theorems and Expected Utilities." *Journal of Economic Theory* 11 (1): 16–34.
- Friedman, Milton, and Leonard J. Savage. 1948. "The Utility Analysis of Choices Involving Risk." *Journal of Political Economy* 56 (4): 279–304.
- Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with Non-Unique Prior." *Journal of Mathematical Economics* 18 (2): 141–53.
- Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics* 112 (2): 631–45.
- Green, Richard C., and Sanjay Srivastava. 1986. "Expected Utility Maximization and Demand Behavior." *Journal of Economic Theory* 38 (2): 313–23.
- Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica* 59 (3): 667–86.
- Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2017. "Non-Parametric Bounds for Non-Convex Preferences." *Journal of Economic Behavior and Organization* 137: 105–12.
- Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2018. "Parametric Recoverability of Preferences." *Journal of Political Economy* 126 (4): 1558–93.
- Harbaugh, William T., Kate Krause, and Timothy R. Berry. 2001. "GARP for Kids: On the Development of Rational Choice Behavior." *American Economic Review* 91 (5): 1539–45.
- Hey, John D., and Chris Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62 (6): 1291–1326.
- Hey, John D., and Noemi Pace. 2014. "The Explanatory and Predictive Power of Non Two-Stage-Probability Theories of Decision Making under Ambiguity." *Journal of Risk and Uncertainty* 49 (1): 1–29.
- Imai, Taisuke, and Colin F. Camerer. 2018. "Estimating Time Preferences from Budget Set Choices Using Optimal Adaptive Design." Unpublished.
- Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review* 97 (4): 1047–73.

- Kubler, Felix, Larry Selden, and Xiao Wei.** 2014. "Asset Demand Based Tests of Expected Utility Maximization." *American Economic Review* 104 (11): 3459–80.
- Loomes, Graham.** 1991. "Evidence of a New Violation of the Independence Axiom." *Journal of Risk and Uncertainty* 4 (1): 91–108.
- Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini.** 2006. "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences." *Econometrica* 74 (6): 1447–98.
- Mattei, Aurelio.** 2000. "Full-Scale Real Tests of Consumer Behavior Using Experimental Data." *Journal of Economic Behavior and Organization* 43 (4): 487–97.
- Nishimura, Hiroki, Efe A. Ok, and John K.-H. Quah.** 2017. "A Comprehensive Approach to Revealed Preference Theory." *American Economic Review* 107 (4): 1239–63.
- Polisson, Matthew, and John K.-H. Quah.** 2013. "Revealed Preference in a Discrete Consumption Space." *American Economic Journal: Microeconomics* 5 (1): 28–34.
- Polisson, Matthew, John K.-H. Quah, and Ludovic Renou.** 2020. "Replication Data for Revealed Preferences over Risk and Uncertainty." American Economic Association [publisher] Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/ExxxxxxV1>.
- Quah, John K.-H.** 2014. "A Test for Weakly Separable Preferences." University of Oxford Department of Economics Discussion Paper 708.
- Quiggin, John.** 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization* 3 (4): 323–43.
- Smeulders, Bart, Frits C. R. Spieksma, Laurens Cherchye, and Bram De Rock.** 2014. "Goodness-of-Fit Measures for Revealed Preference Tests: Complexity Results and Algorithms." *ACM Transactions on Economics and Computation* 2 (1).
- Varian, Hal R.** 1982. "The Nonparametric Approach to Demand Analysis." *Econometrica* 50 (4): 945–73.
- Varian, Hal R.** 1983a. "Non-Parametric Tests of Consumer Behaviour." *Review of Economic Studies* 50 (1): 99–110.
- Varian, Hal R.** 1983b. "Nonparametric Tests of Models of Investor Behavior." *Journal of Financial and Quantitative Analysis* 18 (3): 269–78.
- Varian, Hal R.** 1988. "Estimating Risk Aversion from Arrow-Debreu Portfolio Choice." *Econometrica* 56 (4): 973–79.
- Varian, Hal R.** 1990. "Goodness-of-Fit in Optimizing Models." *Journal of Econometrics* 46 (1–2): 125–40.

AQ1

AUTHOR QUERIES

40

PLEASE ANSWER ALL AUTHOR QUERIES (numbered with “AQ” in the margin of the page). Please disregard all Editor Queries (numbered with “EQ” in the margins). They are reminders for the editorial staff.

| AQ# | Question | Response |
|-----|----------|----------|
|-----|----------|----------|

- | | | |
|----|---|--|
| 1. | Please provide your ICPSR number so we can complete this link to your data. | |
|----|---|--|